CS&SE@SW 2025: a quantitative review of the 8th Workshop for Young Scientists in Computer Science & Software Engineering

Serhiy O. Semerikov[1,2,3,4,5], Andrii M. Striuk[6,1,5]

[1]Kryvyi Rih State Pedagogical University, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine
[2]Institute for Digitalisation of Education of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
[3]Zhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine
[4]Center for Information-analytical and Technical Support of Nuclear Power Facilities Monitoring of the NAS of Ukraine, 34a Palladin Ave., Kyiv, 03142, Ukraine
[5]Academy of Cognitive and Natural Sciences, 54 Universytetskyi Ave., Kryvyi Rih, 50086, Ukraine
[6]Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

Abstract

This paper reviews the 23 papers of CS&SE@SW 2025, comprising the 22 papers of the present volume together with one presented at the workshop and published elsewhere, 305 pages in all. The review is carried out as a measurement exercise: every quantity reported here is produced by a named script from deposited data. Bibliometric structure is measured over the 675 references the volume cites. Topical structure is recovered using a 53-concept controlled vocabulary together with a bootstrap-audited Ward partition of document embeddings, validated against the editors' independent grouping. The recorded discussion is segmented and coded exhaustively, and the volume is positioned against 2024–2026 literature retrieved under a stated protocol admitting no reference without a resolvable DOI. Concept and facet coding was carried out by five language models from five vendors, with agreement reported for every code.

The two structures a bibliometric review conventionally presents are absent on this corpus. The 154 author keywords yield 146 unique tokens, only 6 of which recur, and 4 of 453 unique cited works are cited by more than one paper, giving 3 coupling edges across 253 pairs. The profile of cited *venues* does carry structure, so the negative result is specific to shared handles and not general to bibliometrics. The controlled vocabulary recovers 33 shared handles where the author keywords supplied 6. Applying the same instruments to the discussion surfaces a convergence the corpus alone does not reveal: two papers sharing no author, institution or reference independently adopt schema-constrained tool contracts as the action space for language-model agents, at cosine 0.650 against a corpus median of 0.18. Of the 59 coded questions across 19 discussion rounds, 52 came from a chair and 6 from the floor, with 1 left unattributed.

The review closes with 17 open problems that survive a filter requiring support from at least two of three independent streams of evidence. Reliability is reported for every coded dimension, including one that falls below the level at which it could support a claim.

Keywords

computer science, software engineering, workshop proceedings, bibliometrics, semantic clustering, discourse analysis, large language model agents, research agenda

1. Introduction

1.1. Scope and approach

CS&SE@SW 2025, the 8th Workshop for Young Scientists in Computer Science & Software Engineering, was held online from Kryvyi Rih on 21 November 2025. The workshop belongs to a series that began in 2018 and has appeared in CEUR Workshop Proceedings for seven of its eight editions [1, 2, 3, 4]. Its

CS&SE@SW 2025: 8th Workshop for Young Scientists in Computer Science & Software Engineering, November 21, 2025, Kryvyi Rih, Ukraine
https://doi.org/10.55056/cssesw2025/paper00
semerikov@gmail.com (S. O. Semerikov); andrey.n.stryuk@gmail.com (A. M. Striuk)
https://acnsci.org/semerikov (S. O. Semerikov); http://mpz.knu.edu.ua/andrij-stryuk/ (A. M. Striuk)
0000-0003-0789-0272 (S. O. Semerikov); 0000-0001-9240-1976 (A. M. Striuk)

remit is deliberately broad: it accepts work from doctoral students and early-career researchers across the whole of computer science and software engineering, and it does not organise itself around a single theme.

That breadth is what makes the volume worth measuring, not merely summarising. A collection assembled from whatever early-career work is ready in a given year need not be thematically coherent, and whether it is coherent is an empirical question with a determinate answer. This paper treats it as one. Every claim made here about the volume is a measurement, each measurement names the script that produced it, and the results that came out unfavourably are reported alongside those that did not.

The first thing measurement establishes is that the conventional evidence for thematic coherence is unavailable on a corpus of this kind. The 23 papers supply 154 author keywords, of which 146 are unique, and 675 cited references resolving to 453 distinct works, of which 4 are cited by more than one paper. Drawn as networks, the keyword co-occurrence map and the citation-coupling graph are both fields of isolated nodes (§4.3). The structure that does exist in this volume has to be recovered by other means, and most of what follows is an account of how.

1.2. Contributions and how to read this paper

1. A negative methodological result. Keyword co-occurrence and bibliographic coupling carry no usable structure on a young-scientists workshop corpus, whereas the profile of cited *venues* does. The failure is therefore specific to shared handles and does not generalise to bibliometric methods as a class (§4.3).
2. A method that recovers the structure and reports its own reliability: a 53-concept controlled vocabulary yielding 33 shared handles where author keywords yield 6, and a bootstrap-audited Ward partition of document embeddings, validated against the editors' independent grouping (§2.4, §5.1).
3. A substantive finding that the corpus alone does not reveal. Two papers with no shared author, institution or reference converge on schema-constrained tool contracts as the action space for language-model agents, at cosine 0.650 against a corpus median of 0.18 (§9).
4. A quantitative account of the workshop's recorded discussion: 59 questions across 19 rounds, coded with published reliability (§7).
5. A research agenda derived by triangulation: 17 open problems surviving a filter that requires support from at least two of three independent streams, drawn from 23 candidates, with the 6 discarded single-stream items also reported (§10).

Readers interested primarily in the analytical results should turn to §4.3, §9 and §12. Readers looking for a particular paper should turn to §6, where each of the 23 papers is reviewed and illustrated. Readers intending to apply the same procedure to another volume should read §2 together with §11, which sets out what the design cannot establish.

1.3. Artefacts and reproducibility

Every statistic in this paper is generated. No figure is typed into the prose by hand; each expands a macro written by a named script into `source/gen/`, so that a re-run of the analysis cannot leave the text behind. The scripts, the codebook, the prompts, the coding ledgers, the transcript segmentation with its cue fingerprints, the reliability computations and all derived tables are deposited at https://github.com/ssemerikov/cssesw2025-paper00-artefacts.

Three items are deliberately withheld, for the same reasons that they do not appear in this paper. The discussion transcript is not published: it is a 32574-word automatic transcript of identifiable early-career researchers speaking, so only the derived, role-labelled coding is released. The mapping from reviewer pseudonyms to names is not published, and no reviewer is named or quoted anywhere in this paper, since in a committee of this size a single quoted sentence would identify its author. The credentials for the retrieval tooling are held outside the repository.

2. Methodology

This section states what was measured, from which artefact, under which rule, and how reliably. It is set out at length because the claims made later are measurements, and a measurement that cannot be re-derived carries no more weight than an impression. Everything reported here is reproducible from the artefact bundle described in §1.3: each figure in the prose is a macro written by a named script, not a literal typed into the source.

2.1. Corpus definition, inclusion criteria and provenance

The corpus comprises the 22 papers accepted into this volume together with one further paper, Fedorchenko et al. [5], presented at the workshop and published elsewhere, giving 23 reviewed papers across 305 pages. The unit of analysis is the workshop, and the delivered programme contained that talk; excluding it would misdescribe the event. It is cited throughout to its published record, under the title it carries there, which differs from the title on its CEURART source. The discrepancy is noted where it matters.

Provenance is narrow by design. Paper content is taken from the camera-ready `paperNN.tex` sources and the PDFs built from them, and never from a secondary description. A paper's title is taken from its own source alone. Screenshot filenames, the authors' registry export, the CEUR index and, for Fedorchenko et al. [5], the journal's own record all exist and can disagree, so `analysis/check_titles.py` fails the build if a title used here drifts from the one in the paper it describes. Bibliometric measurements read the `.bbl` files, which are machine-readable item by item, so the population measured is what each paper actually cited rather than whatever its `.bib` file contained. The two differ by 31 entries, and both counts are reported.

Two rules govern the bibliography of this paper. First, a reference is admitted only if it has a resolvable DOI, and every entry is rebuilt from that DOI by content negotiation rather than transcribed from a search result or a model answer; the exception is a curated bibliography where no DOI exists, which arises because CEUR-WS volumes are indexed by URL. Second, members of the programme committee are cited under a stated quota, not by selection: one publication for serving, and one further publication for each review filed. The rule follows the previous edition [4] and is made proportional here so that reviewing effort is visible in the bibliography.

2.2. The submission and review record

The review export and the committee roster were supplied by the chairs, and every figure quoted in §3.2 is a primary-source count parsed by script: 92 reviews filed by 23 reviewers over 32 submissions, a mean of 2.88 reviews per submission with a range of 1–4, and ten ordinal criteria per review. 42 submissions were received in total, of which 39 appear as records in the export; the difference consists of incomplete submissions, duplicates and withdrawals, and both counts are given so that the acceptance rate can be read against either denominator.

No reviewer is named anywhere in this paper and no review prose is quoted at any length, because in a committee and a volume of these sizes a single quoted sentence would identify its author. Scores are analysed under pseudonyms and the mapping is not published.

2.3. Concept coding: controlled vocabulary and reliability

Because author keywords cannot carry the volume's structure, for reasons measured in §4.3, a controlled vocabulary was constructed in their place. It contains 53 concepts, drawn from the ACM Computing Classification System wherever that scheme supplies a term and extended with thirteen terms it lacks for this corpus, among them *UAV platform*, *tool-calling contract* and *wartime and demining application*. Eight single-label facets accompany the concepts, covering domain, technique, modality, evaluation strategy, dataset provenance, artefact availability, deployment target and maturity. Each paper was coded from an identical dossier, so that every rater saw the same evidence.

The raters are 5 language models from five vendors, sharing neither weights nor training pipeline, each supplied with the same dossier, codebook and instruction at temperature zero. There was no human coding pass, and the consequence is stated here because it bounds every agreement figure in this paper: agreement among five independent models measures the model-independence of the coding instead of its human reliability, and it cannot establish that the codebook is correct. Five models may share a bias that no human would have, and this design would not detect it. The statistic is accordingly reported as *inter-model agreement* throughout and never as inter-coder reliability, and the absence of a human pass is carried in §11 as a limitation.

Agreement is reported both as Krippendorff's α [6, 7] with a bootstrap confidence interval [8] and as Gwet's AC_1 [9]. The reason for reporting both is visible in the data. Concept coding reaches $\alpha = 0.73$ $[0.69, 0.77]$ with $\mathrm{AC}_1 = 0.94$ and 89.0 % of cells unanimous across all five raters. Artefact availability, where 21 of 23 papers fall into a single category, instead collapses to $\alpha = 0.10$ while AC_1 reads 0.63 and half the papers are coded unanimously. This is the prevalence paradox [10] in a real table: α taken alone would dismiss the coding as noise, while AC_1 taken alone would conceal that almost all of the agreement rests on one dominant category. A second facet, maturity, is weak rather than skewed, at $\alpha = 0.41$, where the models disagree about substance; §5.8 prints that reliability on the figure itself.

The use of language models as annotators is a contested practice with a literature on both sides [11, 12]. The position taken here is that it is acceptable to the extent that it is audited and its failures are reported, of which the artefact facet above is one.

2.4. Semantic cartography: representation, clustering, stability

The volume's topical structure is recovered semantically. Three document units were constructed, comprising title and keywords, the same with the abstract added, and the same again with the section skeleton added. Each was crossed with two representations that share no components: a sparse term-frequency model fitted on these 23 documents, and dense sentence-embedding vectors [13] from a pinned revision of MPNet [14]. A document vector is the normalised mean of its *sentence* vectors, not an encoding of the concatenated text, since the model truncates at 384 word pieces and several abstracts exceed that length; encoding the whole string would have discarded their tails without warning.

Clustering uses Ward linkage on cosine distance [15], cut at $k = 6$, with the cophenetic correlation reported [16]. Neither manifold learning nor density-based clustering was used, because at $n = 23$ neither is defensible; §11 records what else was tried and abandoned. Stability was assessed over 1000 bootstrap draws resampling sentences within each document and is reported as a co-assignment probability of 0.67 within clusters against 0.06 between them, which establishes that the partition does not depend on the particular sentences a dossier happens to contain. Agreement with the editors' own six-way partition is reported as the adjusted Rand index [17] and normalised mutual information, at 0.34 and 0.64 for the reported run; the disagreements are interpreted in §5.1.

One methodological observation may be useful to anyone repeating this procedure on another volume. The two representations agree only moderately, and their agreement increases with the amount of each paper that is read: the adjusted Rand index between them rises from 0.22 on titles and keywords to 0.25 on titles and abstracts, and to 0.54 once section skeletons are included. On a corpus of this size the choice of representation matters appreciably, and reporting any single partition as though it were the structure would overstate what these runs support.

2.5. Discussion-transcript protocol

The recorded discussion is the second body of evidence used here, and the more difficult of the two. The source is a single automatic-captioning transcript of 2657 cues and 32574 words, running just under six hours, carrying no speaker labels and badly garbling names, titles and acronyms. The protocol is built around those two properties.

Attribution is by role and never by name. Utterances are labelled as chair, presenter or floor, and no individual is named on the basis of garbled captioning. The constraint is partly methodological and

mainly ethical, since the presenters are early-career researchers and the authors of this paper chaired the session. It is also consistent with the design, in which the unit of analysis is the question instead of the person; that choice removes the missing-labels problem instead of working around it.

The unit of analysis is an *ask*: a single contiguous move that requests information, challenges a claim or recommends a change, with multi-part questions split into their components. Nineteen of the 23 papers had a live discussion, one having been pre-recorded and one presented without its authors, and all 19 spans were read exhaustively, yielding 59 asks across 51 turns. Exhaustive reading was necessary rather than merely thorough: the automatic pre-filter has a measured recall of 0.56, and three complete rounds produced no candidate at all while containing four asks between them. That recall figure is recomputed on every run.

Quotation follows three declared tiers: verbatim with a timestamp; lightly repaired, where a captioning error is corrected within square brackets and disfluencies are removed silently; and paraphrase with a timestamp span where the text is too corrupted to quote. Content words are never repaired. Where a number, a model name or a metric is corrupted, it is marked unintelligible or the claim is dropped, since inferring what a presenter intended to say is the error the protocol exists to prevent. Every quotation carries a timestamp so that a reader with access to the recording can verify it. Answers are characterised by informational content alone and never by the fluency of the speaker's English, most presenters having worked in a second language.

Reliability was measured on the same design as the concept coding, with four raters over the fixed set of 59 asks. It is not uniform across dimensions:

- what was *asked* is codable: act $\alpha = 0.46$ $[0.34, 0.59]$, topic $\alpha = 0.51$ pooled over binary passes with $\mathrm{AC}_1 = 0.88$;
- what the answer *did* is not: response $\alpha = 0.37$, with only 20 % of asks coded unanimously.

§7.4 accordingly reports the response distribution descriptively and does not treat it as a measurement. Unitisation was not measured at all, the unit set having been established manually and held fixed for every rater, and §11 records that limitation.

2.6. External-literature positioning protocol

Positioning the volume against current work requires an auditable retrieval step, so the procedure is scripted end to end. 26 questions, covering the cross-cutting trend, each thematic cluster and the review's own method, were run unattended against a literature-retrieval assistant, and a deep-research run against the same tool was parsed alongside them. The 27 answers yield 522 reference records: 33 of them carry no DOI and were excluded and logged under the house rule, leaving 489 distinct DOIs. Shortlisting follows a stated rule rather than a global frequency cut: the three references per question on which that answer's prose most often relies, together with anything returned by three or more questions. A global cut would over-represent whichever questions produced the longest answers and would leave some clusters unevidenced, whereas the per-question rule ensures that every cluster and every question is represented. The rule shortlists 77 records, of which 4 were already cited within the volume and 73 were admitted to `external.bib`, each rebuilt from its DOI. Those on which the argument of this paper rests are cited in §8 and §9; the remainder are deposited with the artefact repository, so that the whole shortlist produced by the protocol is available and not only the part quoted here.

The last of those counts is itself a result, and §8 develops it: only 4 of the 77 shortlisted frontier references appear anywhere among the volume's 453 unique cited works, an overlap of roughly 5 per cent on the questions this review asks.

3. CS&SE@SW 2025 at a glance

3.1. The series 2018–2025 and its scope

This is the eighth edition of a workshop held annually since 2018 [18, 19, 1, 2, 20, 3, 4]. The series' record shows no simple growth trend. Acceptance has not moved monotonically: 76 % in 2018, then 57, 60, 68 and 58 %, a sharp tightening to 40 % in 2023, 44 % in 2024, and 52 % in the present edition. Submissions grew more than fourfold between 2020 and a peak of 64 in 2024, then fell back to 42 in 2025. Review depth peaked in 2021 at 3.14 reviews per submission and was thinnest in 2022 at 2.2, with this edition at 2.88. Table 1 and figure 1 give the full series.

Table 1
The eight editions of CS&SE@SW. Submission and acceptance figures are quoted from each edition's own preface or foreword, not from repository metadata; review depth is the mean number of reviews per submission where the edition reported it. Acceptance is not monotone, and the length of the opening paper has varied widely across the series.

Ed.	Year	Proceedings	Subs.	Acc.	Rate	Rev./sub.	Op. pp.	Op. words
1	2018	Vol-2292	25	19	76 %	n/s	10	3 292
2	2019	Vol-2546	30	17	57 %	2.70	20	4 933
3	2020	Vol-2832	15	9	60 %	2.70	10	2 473
4	2021	Vol-3077	22	15	68 %	3.14	35	8 635
5	2022	SCITEPRESS	19	11	58 %	2.20	1	–
6	2023	Vol-3662	42	17	40 %	n/s	36	9 268
7	2024	Vol-3917	64	28	44 %	n/s	46	14 115
8	2025	Vol-3XXX	42	22	52 %	2.88	–	–

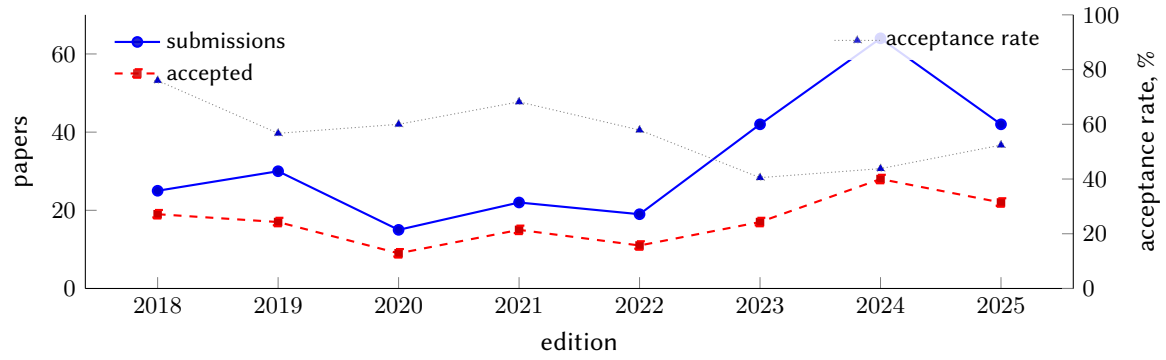


Figure 1: Eight editions of CS&SE@SW: submissions, papers accepted and the acceptance rate. Acceptance is not monotone, selectivity tightened sharply at the sixth edition and has since relaxed, and the submission peak of 2024 is not sustained. Figures are quoted from each edition's own preface (table 1).

One feature of the series bears directly on §7: discussion time has been compressed as the workshop has grown. The first edition allowed fifteen minutes of presentation and seventeen minutes of discussion per paper, whereas the 2025 acceptance letter allocates seven to ten minutes of talk and five to seven of discussion. Discussion has thus moved from being longer than the talk to roughly half its length.

3.2. Submissions, review load and acceptance

The workshop received 42 submissions and accepted 22 into the volume, a rate of 52 %, with one further paper presented and published elsewhere [5]. Of the 39 submissions that appear as records in the review export, 32 carry visible reviews; 12 were rejected after review and 4 were desk-rejected without going to a reviewer, which is why the two are counted separately here, not added into a single rejection rate.

Across ten ordinal criteria the reviewers were generous about relevance, with a mean of 4.03, and about presentation, but strict about novelty, which averaged 2.43 against 3.3–4.0 on every other criterion. The highest value on the scale was awarded four times across 92 reviews and the lowest nine times. Whether this reflects calibration or an accurate reading of early-career work is a question the volume can pose but not settle; novelty is nonetheless the one criterion on which this committee did not hedge. The overall score separates accepted from rejected submissions cleanly, at 3.65 against 2.79.

3.3. The programme as planned and as delivered

The delivered programme can be recovered independently of the transcript from the numbering of the presentation screenshots, and the agreement of those two signals is what makes the reconstruction in §7 defensible. Comparing the planned running order with what took place gives figure 2. The order largely held: two talks exchanged slots, one paper's authors were absent at the start so their slides were shown at the end, and one paper was played from a recording in the penultimate slot.

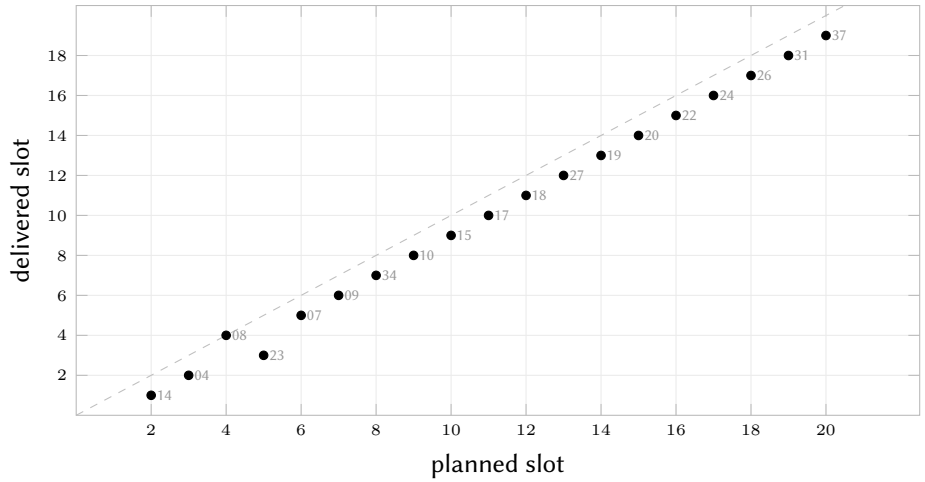


Figure 2: The programme as planned against the programme as delivered. The planned order is the numbering of the chairs' own prepared question sets; the delivered order is reconstructed from two independent signals, the chair handover formulas in the recording and the numbering of the presentation screenshots. Marks are labelled with the paper number. The order largely held: papers 23 and 08 exchanged slots, and the one paper whose authors were absent at the start fell out of slot 1 altogether and was shown at the end.

The delivered schedule departed substantially from the planned one. Against an allocation of seven to ten minutes, talks ran from 5.5 to 24.1 minutes with a mean of 11.4, and only five of the 19 rounds fell inside the allocated window; the longest talk ran to almost two and a half times its allocation. Discussion ran from 1.8 to 11.5 minutes against an allocation of five to seven, with six of 19 inside it. Substantive talks were allowed to overrun and the time was recovered from discussion, a trade-off taken up in §7.

3.4. Programme committee

31 of the committee members announced for the workshop appear on its review record, and 17 of those filed at least one review. The list below follows the review record rather than the announcement: announced members who do not appear on that record are not listed. Under the rule stated in §2.1, each member is cited once for serving and once more for each review filed.

- *Nadire Cavus*, Near East University, Cyprus [21]
- *Stuart Charters*, Lincoln University, NZ [22]
- *Roman Danel*, Institute of Technology and Business in České Budějovice, CZ [23, 24, 25, 26]
- *Emre Erturk*, Eastern Institute of Technology, NZ [27]

- *Helena Fidlerová*, Slovak University of Technology, SK [28, 29, 30, 31]
- *Oleksii Haluza*, National Technical University "Kharkiv Polytechnic Institute", UA [32, 33, 34, 35, 36]
- *Pavlo Hryhoruk*, Khmelnytskyi National University, Ukraine [37]
- *Dragos-Daniel Iordache*, National Institute for Research and Development in Informatics - ICI Bucuresti, RO [38]
- *Oleksandr Kolgatin*, Simon Kuznets Kharkiv National University of Economics, Ukraine [39]
- *Valerii Kontsedailo*, Inner Circle, Netherlands [40]
- *Hennadiy Kravtsov*, Kherson State University, UA [41, 42, 43, 44, 45]
- *Vyacheslav Kryzhanivskyy*, R&D Seco Tools AB, Sweden [46]
- *Nagender Kumar*, India [47, 48, 49]
- *Andrey Kupin*, Kryvyi Rih National University, Ukraine [50, 51, 52, 53, 54, 55, 56, 57, 58]
- *Nadiia Lobanchykova*, Zhytomyr Polytechnic, UA [59, 60, 61, 62]
- *Orken Mamyrbaev*, Institute of Information and Computational Technologies, Kazakhstan [63]
- *Mykhailo Medvediev*, ADA University, AZ [64, 65]
- *Bongkyo Moon*, QIR, Korea, South – Republic of Korea [66, 67, 68, 69]
- *Michael O'Grady*, University College Dublin, Ireland [70]
- *Vasyl Oleksiuk*, Ternopil Volodymyr Hnatiuk National Pedagogical University, UA [71, 72, 73, 74, 75]
- *Jaderick P. Pabico*, University of the Philippines Los Baños, PH [76, 77, 78, 79, 80]
- *Grażyna Paliwoda-Pękosz*, Krakow University of Economics, Poland [81]
- *James Procter*, University of Dundee, United Kingdom [82]
- *Oleg Pursky*, Kyiv National University of Trade and Economics, UA [83]
- *Serhiy Semerikov*, Kryvyi Rih State Pedagogical University, UA [84]
- *Etibar Seyidzade*, Azerbaijan Technical University, AZ [85, 86, 87]
- *Andrii Striuk*, Kryvyi Rih National University, Ukraine [88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 43, 99, 100]
- *Tetiana Vakaliuk*, Zhytomyr Polytecnic State University, Ukraine [101, 102, 103, 104, 105, 62, 106]
- *Nataliia Veretennikova*, Lviv Polytechnic National University, Ukraine [107, 108, 109, 110]
- *Volodymyr Voytenko*, Athabasca University, Canada [111, 112]
- *Alejandro Zunino*, ISISTAN - UNCPBA & CONICET, AR [113, 114, 115, 116, 117]

A further 6 people reviewed for the workshop without appearing on the announced committee, filing 22 reviews between them. They are listed here as additional reviewers, under the same citation rule.

- *Christopher Anand*, McMaster University, Canada [118, 119, 120, 121, 122]
- *Shiva Kumar Bhuram*, Dorman Products Inc., USA, US [123, 124, 125, 126]
- *Suresh Dameruupula*, MYGO Consulting Inc, US
- *Andriy Dudnik*, Taras Shevchenko National University of Kyiv, UA [127, 128, 129, 130, 131, 132]
- *Praveen Ellupai Asthagiri*, Amazon, US
- *Surya Bhaskar Reddy Karri*, Pinterest, US

Together the two groups filed 92 reviews. Where a member's indexed record contains fewer publications than the rule allots, the entries that exist are cited and no substitute is supplied.

3.5. Organisers

The workshop is organised by the Academy of Cognitive and Natural Sciences together with the Department of Simulation and Software Engineering at Kryvyi Rih National University, and was held online on 21 November 2025. CEUR-WS.org supports the event as well as publishing it: its editorial work carries these proceedings into the record, as it has done for seven of the eight editions listed in table 1.

The online format is not a consequence of the pandemic. The series was hybrid by design from its first edition, which in 2018 was streamed with questions taken from remote participants. In 2025 it is also a wartime necessity: the session was scheduled around the possibility of power failures, and the opening remarks say so.

4. Bibliometric and structural profile of the corpus

4.1. Corpus inventory and authorship geography

The 23 reviewed papers run to 305 pages and carry 87 author slots, 81 of them in the 22 papers of this volume, giving a mean of 3.8 authors per paper across 37 distinct affiliation strings. The spread behind that mean is wide: papers range from 8 to 21 pages, from 0 to 38 figures, and from 6 to 167 supplied references. Any per-paper average in this section should be read with that variation in mind.

The geography is concentrated but not closed. Twenty-two of the 23 papers carry at least one Ukrainian affiliation, one is wholly Vietnamese, and international co-authors appear from the United Kingdom, Spain, Indonesia and Poland. The volume is therefore best described as a national workshop with an international edge. The conditions of its production are visible in the programme as well as the affiliations: the session was held online and scheduled around the possibility of power failures. Several papers take that context as their subject matter, among them work on demining, airspace monitoring and resource-constrained computing, and §5 reports it as a facet of the corpus.

4.2. The reference base

Two reference populations are reported separately because they differ and the difference is informative: 706 entries were supplied in the papers' bibliography databases, while 675 were actually cited. All structural claims below use the cited population, read from the typeset bibliographies instead of the databases, so that an unused database entry cannot inflate a count. Of the cited references, 457 carry a DOI, resolving to 453 distinct works.

The base is unusually current. 352 of the dated cited references, 52 %, are from 2024 or later, and the modal year is 2024 (figure 3); more references are dated 2026 than 2020 and 2021 combined. For a volume written largely by early-career researchers this runs against the expected pattern.

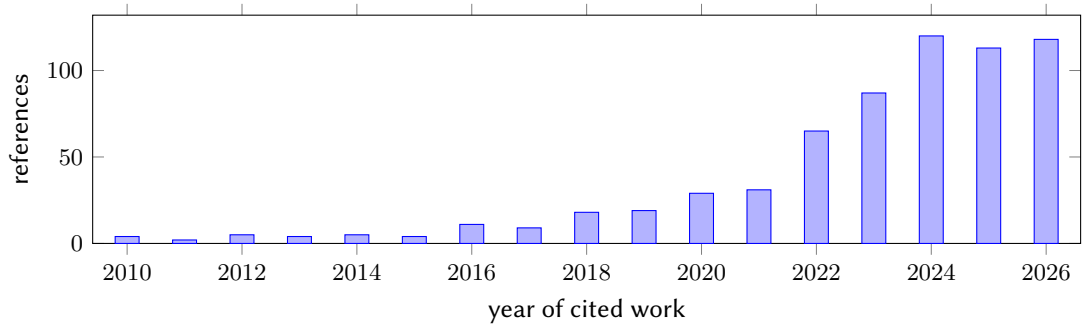


Figure 3: Year of the 675 references cited across the volume. 52 % are from 2024 or later and the modal year is 2024; more references are dated 2026 than 2020 and 2021 together. One cited reference carries an impossible year, 2028, and is left in the figure rather than quietly dropped.

Currency carries a cost, visible in figure 4: 117 of the 706 supplied entries, 17 %, are typed as miscellaneous, so grey literature, preprints and web resources make up a sixth of the evidence base. Currency and peer-review status trade against each other, and a volume citing the 2026 literature heavily necessarily cites much of it before review. The tension is not peculiar to this volume: preprint discovery, citation and policy remain contested across 2024–2026, with no settled rule for what a preprint citation warrants [133, 134, 135]. The most-cited venues are otherwise conventional and sensor- and remote-sensing-heavy: *Sensors* (8 citing papers), *IEEE Access* (7), *Remote Sensing of Environment* and *ACM Computing Surveys* (6 each).

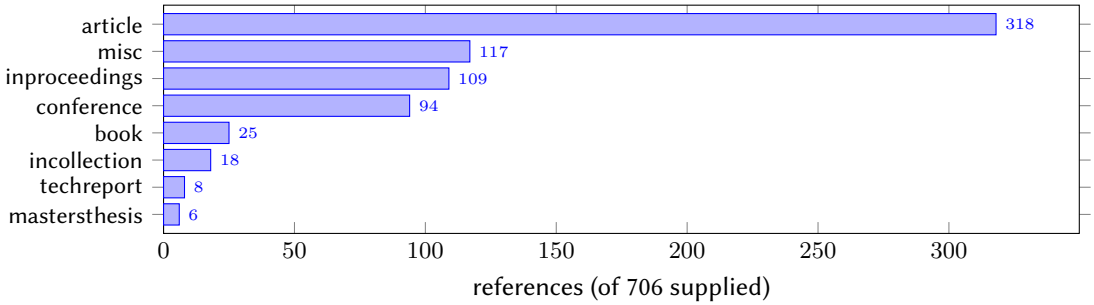


Figure 4: Entry types across the 706 supplied bibliography records. 117 (17 %) are miscellaneous (preprints, reports and web resources) so grey literature is a sixth of the evidence base. This is the cost side of the currency reported in figure 3.

4.3. Conventional bibliometric structure is absent

A review of a proceedings volume conventionally presents two figures: a keyword co-occurrence map and a citation-coupling network [136, 137, 138], usually drawn with a standard tool [139] and following a standard recipe [140]. On this corpus both are empty, and the emptiness is itself the result.

The measurements are in table 2, over the 253 possible pairs of 23 papers. Author keywords supply 154 tokens of which 146 are unique, only 6 recurring at all, giving 9 edges, a density of 0.036, a largest component of 5 papers and 12 isolates. Bibliographic coupling is sparser still: of the 453 unique cited DOIs, 4 are cited by more than one paper, and since one pair of papers shares two of them these yield 3 edges across 253 pairs, a density of 0.012, a largest component of 2 and 17 isolates. Drawn, either map is a field of isolated points.

Table 2
Conventional bibliometric structures on this corpus, and the semantic graph compared at the same edge count rather than at an arbitrary similarity threshold. Two of the three structures a bibliometric reader expects are all but empty, which is the finding that motivates the semantic cartography rather than an excuse for it.

structure	nodes	edges	density	largest	isolates	verdict
author keyword sharing	23	9	0.036	5	12	a field of isolates: 12 of 23 papers share no keyword with any other
bibliographic coupling	23	3	0.012	2	17	3 edges of 253 possible pairs; 17 isolates
cited-venue profile	23	38	0.150	16	7	the one conventional signal with real structure
semantic cosine (TA/dense)	23	9	0.036	—	—	matched to the keyword graph's edge count

The result is specific, not general, and the distinction matters. A third conventional signal, the profile of *venues* each paper cites, does carry structure: 38 edges, a density of 0.150 and a largest component of 16 papers. The claim is therefore not that bibliometrics fails on small corpora. It is narrower: on a young-scientists workshop volume the two structures built from shared *handles*, namely the keywords

the authors chose and the particular works they cited, are empty, while the structure built from where the field publishes survives. Early-career authors writing across 23 largely unrelated problems share a literature at the level of venues and almost never at the level of individual papers.

Two consequences follow. An assertion of thematic coherence for this volume would assert something the corpus does not contain, and the standard method would not detect the absence, because the graph is not usually drawn before the claim is written. The structure must therefore be recovered by other means, which is the business of the rest of this section. One further method was tried and abandoned: a technology-radar construction built from citation contexts, which this corpus cannot support, for reasons given in §11.

4.4. Semantic cartography of the corpus

The first replacement is lexical but controlled. Coding all 23 papers against the 53-concept vocabulary of §2.3 uses 52 of the concepts and yields 33 concepts carried by more than one paper, against 6 for the authors' own keywords, or 5.5 times as many shared handles. The matrix remains sparse in absolute terms, at 139 of 1219 cells or 11.4 % dense, with between 3 and 11 concepts per paper and a mean of 6.0. The two observations belong together: a controlled vocabulary does not make a heterogeneous volume homogeneous, but it does make the overlaps that exist visible.

The second replacement is semantic, and it is what §5 is built on. Ward clustering of dense document embeddings, cut at $k = 6$, gives a partition with a cophenetic correlation of 0.60 that agrees with the editors' independent six-way grouping at an adjusted Rand index of 0.34 and normalised mutual information of 0.64. Under 1000 bootstrap draws resampling sentences within each document, mean co-assignment is 0.67 within cluster against 0.06 between, the partition survives resampling of the very text it is computed from.

Four pairs of papers co-cluster in every one of the six representation-and-unit runs, and two of those pairs recur later in this paper. The most similar pair in the volume, at cosine 0.662, crosses the editors' own boundary: a game-engine performance study and a visual-novel history of software engineering are textually the same kind of paper, a disagreement discussed in §5.1 and not resolved by reassignment. The second-ranked pair, at cosine 0.650 against a median of 0.18 over all 253 pairs, is the convergence examined in §9. Those two papers share no reference and no keyword, which is why neither map in §4.3 can detect them.

5. Thematic synthesis: six clusters

5.1. How the clusters were derived and validated

Two partitions of the same 23 papers exist and both are reported here. The editors' partition was recorded before any clustering was run, on the basis of reading the papers: six groups with exclusive membership, each named for a research problem rather than a technique. The data-driven partition is the Ward cut of §4.4. The two agree substantially but not identically, at ARI $= 0.34$ and NMI $= 0.64$, and table 3 shows where they diverge.

The reported run recovers one a-priori cluster exactly and divides another. All four airborne-sensing papers cluster together, the only editors' group to survive intact. The largest a-priori cluster does not survive: its six papers separate cleanly into three concerning signals and dynamics [141, 142, 143] and three concerning systems with an adversary or a protocol [144, 145, 146], the latter joining the cluster that contains the two language-model papers. The volume's largest editorial group is thus better described as two groups, and the measurement shows this more clearly than the reading did.

Three further crossings are reported because each is a substantive claim about the corpus, not noise. A game-engine performance study and a visual-novel history of software engineering [147, 148] form a pair at cosine 0.662, the highest of all 253 pairs, across the editors' boundary between interactive graphics and disciplinary reflection: both are, textually, reports of building an interactive artefact and measuring or narrating what it does. A study of VBA macros in science teaching pairs with a study of

Table 3
The editors' six a-priori clusters (rows) against the six-way Ward partition of the document embeddings (columns), for the reported run. Agreement is ARI $= 0.34$ and NMI $= 0.64$: substantial but far from identity. Every off-diagonal entry is discussed in §5.1 rather than resolved by moving a paper.

	k_1	k_2	k_3	k_4	k_5	k_6	n
C1 *language models as substrate*	–	2	–	–	–	2	4
C2 *airborne sensing*	–	–	–	–	4	–	4
C3 *applied ML*	–	–	–	1	1	2	4
C4 *signals and security*	3	3	–	–	–	–	6
C5 *interactive graphics*	–	–	2	–	–	1	3
C6 *computing education*	–	–	1	1	–	–	2
cluster size	3	5	3	2	5	5	23

motivation in inclusive workplaces [149, 150] in five of six runs: both are questionnaire-and-expert-method studies of practitioners, which a domain-based partition separates and a method-based one does not. And an air-pollution risk system [151] sits with the remote-sensing papers and not with applied machine learning, because the cartography reads method and the editors read purpose.

The design for this analysis predicted which papers would prove to be boundary cases, and the prediction was not borne out: it named the facial-authentication paper and the simulator meta-review, whereas the measured boundary cases are the forest-cover study and the visual-novel study. No paper was reassigned in order to improve the agreement index. The exclusive membership used below is the editors' partition, with crossings reported where they bear on the argument.

5.2. Language models as software substrate

Four papers [152, 153, 154, 155] treat a language model not as an object of study but as a component with an interface. Two of them make the same architectural bet independently and are the subject of §9: a case for GraphQL as a typed, schema-validated action space for agents, with a token-economy analysis (63 request tokens down to 32; 245 operational tokens down to 175) and a security argument that a schema-bounded language is a native sandbox [152]; and a conversational e-commerce assistant in which a model emits Cypher against a Neo4j product graph, with Model Context Protocol servers enforcing validation so the database is never directly exposed [153].

The other two are about what models do to a method. A sociological NLP study finds that 107 of 990 Ukrainian news headlines (10.8 %) change their sentiment polarity according to nothing but the choice between lemmatisation and stemming [154]: preprocessing as a substantive methodological decision rather than a technical preliminary, which is a result this paper's own §2.4 had to take seriously. A systematic review of text generation [155] covers 43 Scopus articles from 2022–2024 and then, unusually, adds a declared non-systematic section on 2024–2026 because its search window had closed: Transformer dominance persisted, but the locus of innovation moved to post-training, inference-time compute, retrieval and evaluation. That is a review explicit about its own expiry date, and its conclusion points the same way as §9.

5.3. Airborne sensing and geospatial intelligence

The cluster the cartography recovers exactly [156, 157, 158, 159] is also the one where the wartime context is least separable from the engineering. Three papers are operational and one is a meta-review of the tools that stand in for reality.

The operational three span the sensing chain. A UAV-and-infrared system for detecting and mapping explosive remnants of war [156] works the optics through to numbers (0.12 m object diameter, 150 mm focal length, five pixels needed for detection), calibrates the camera, projects image points with OpenCV, and reports that error grows away from the optical axis and that automatic exposure costs contrast near

the ground. A multi-sensor forest-change study [158] combines Sentinel-1 SAR, Sentinel-2 multispectral and a DEM over the Skole Beskids and compares a random forest against a UNet with a ResNet-34 backbone; the deep model wins on boundary delineation (IoU 0.84, F_1 0.91, pixel accuracy 0.96) and the study locates 1.247 km^2 of change, some of it in military forest districts. A real-time counter-UAV pipeline [159] chains YOLOv12 detection, BYTETrack identity maintenance and a position-coded Transformer for short-horizon trajectory forecasting, reporting RMSE 12.66 pixels, MOTA 0.79, IDF1 0.64, mAP_{50} 91.04 % on 593,802 Anti-UAV frames, at 55 FPS on GPU and 13 on CPU.

The fourth [157] is the cluster's methodological conscience and the volume's clearest example of a secondary study earning its place. It consolidates nine comparisons of UAV simulators from 2018–2025 and finds that they share a vocabulary of evaluation axes but neither name nor weight them alike, so a platform's standing depends on which survey you read; that coverage is grossly uneven, Gazebo assessed five times and most platforms once; and that three of seven platforms in one comparison have no documentation newer than 2016. It concludes that the field's gap lies not in simulator capability but in the absence of a shared benchmarking protocol, and specifies what such a protocol would have to fix, including the reporting of measured sim-to-real discrepancy, which no reviewed study does reproducibly.

5.4. Applied machine learning for health, environment and industry

Four papers [5, 151, 150, 160] point a learned model at a decision a human would otherwise make, with an expert somewhere in the loop. Two are model bake-offs done properly. An air-quality system [151] trains six models (decision tree, SVM, random forest, XGBoost, CNN, LSTM) on identically preprocessed data and compares them at the evaluation stage instead of chaining them, reaching 0.988 accuracy with the random forest against 0.856 for the single tree, and then makes the argument that decides the design: training time ranges from 3 s to 67 s, so the random forest is the better engineering choice, not merely the more accurate model. A textile-composition classifier [160] puts ViT-B/16 against EfficientNet-B0 and ConvNeXt-Tiny on fabric macrostructure images and reports 98.2 % accuracy for the transformer, positioning it explicitly as an alternative to expensive spectral analysis, and it is one of only 2 papers in the volume that release a dataset.

The other two are about people. A diet-design study [5] compares four optimisation algorithms and three predictors on accuracy, speed and the number of distinct diets produced, which is an unusually sensible third criterion. A motivation study in inclusive education [150] builds an expert system over questionnaire data; it is the cluster's methodological outlier, and the cartography pairs it with the VBA teaching paper and not with its domain neighbours for exactly that reason.

5.5. Signals, security and resource-constrained computing

Six papers [141, 144, 145, 146, 142, 143] form the volume's centre of gravity, and the fact that the cartography splits them is the most useful thing in table 3. The signals half is classical and quantitative: a modified steering vector making a generalized sidelobe canceller robust to head movement and microphone mismatch, improving SNR from 7.4 to 8.8 dB [141]; and a method for designing a chaotic system with a *predefined* attractor by differentiating an algebraic attractor equation, realised as a cardioid-like attractor on an Arduino Due and reporting hidden chaos [142].

The systems half is bounded by an adversary or a protocol. A face-authentication module [144] uses a ResNet-34 embedding with MediaPipe FaceMesh liveness checks, reporting 96.8 % on LFW and CelebA. A post-quantum implementation [146] builds Kyber from the MLWE problem as a .NET library, exercises the full keygen–encapsulate–decapsulate cycle, benchmarks the parameter sets single- and multi-threaded, and gets it running on mobile through .NET MAUI. A simulation study [145] compares MQTT, CoAP and HTTP under IoTSim-Edge and finds CoAP 40.6 % faster than MQTT (22.7 ms against 42.0) at 9.9 % less energy, while HTTP's ≈4.91 s response time rules it out of real-time use.

The sixth [143] belongs to both halves and to §9's secondary theme: a review arguing neuromorphic spiking networks into cognitive radio for sub-millisecond spectrum decisions on Loihi, TrueNorth

and SpiNNaker, claiming two orders of magnitude in energy efficiency. Read against the other two latency-bound papers here and in §5.3, three papers arrive at a hard latency budget from three unrelated directions: spiking hardware, real-time detection, and edge-protocol choice.

5.6. Interactive graphics, virtual reality and perceptual interfaces

Three papers [147, 161, 162] treat rendering, embodiment and colour perception as engineering constraints with measurable costs. One measures what a game engine costs in frames per second and GPU and CPU utilisation across Unity, Unreal, Godot and Redot, building a 3D platformer in Unreal 5.4.4 with Blueprint to do it [147]. One specifies a reproducible pipeline from Blender rigging to Unity XR Interaction Toolkit, formalising a VR scene as agents, interactive objects and trigger zones driven by a finite-state machine over baked clips [162]. The third is the volume's quietest paper and arguably its most transferable [161]: it applies colour-difference formulae and discrimination thresholds to the choice of interface colours, including for users with colour-vision deficiency, and compares its metric against contrast ratio. It is a single-author paper that produces a usable rule, and it drew the shortest discussion of the workshop (§7).

5.7. Computing education and disciplinary reflection

Two papers [149, 148] have the discipline teach and narrate itself. One reports a teacher-training programme built on VBA macros for didactic visualisation in science and mathematics, and its finding is that the approach works *for teachers with no prior programming experience*, which is why it is one of only 2 papers in the volume to reach the second-highest maturity level. The other builds a Ren'Py visual novel about the 1968 NATO Software Engineering conference, taking its characters and dialogue from the conference report itself and generating locations and portraits from period photographs with AI. It is the volume's most explicit act of disciplinary self-reflection, and it is also, as §5.1 records, textually the twin of the game-engine performance study, which no keyword either paper chose would ever have revealed.

5.8. Cross-cutting facets: evaluation, data, deployment, maturity

Table 4 gives the full facet matrix. Four readings of it matter, and one of them is favourable in a way prefaces rarely get to claim.

Every paper in the volume evaluates something. The "no evaluation" category of the evaluation facet is used 0 times in 23 papers, and the modal strategy (7 papers) is a controlled comparison of alternatives on the authors' own data. For a young-scientists workshop where reviewers scored novelty at 2.43 (§3.2), the implication is that the weakness the committee identified is not a want of empirical discipline.

Almost nothing is released. 21 of 23 papers make no artefact available; 2 do. One further paper cites a Zenodo package, but it is a third party's library rather than the paper's own contribution, so it does not count under the rule. This is the facet where the reliability statistics matter most, since one category holds 21 of 23 cases drives α to 0.10 while AC_1 reads 0.63 (§2.3), and it is also, given the mistyped repository URL in §4.2, the volume's most correctable weakness.

Nothing is deployed. Maturity reaches 4 at most, never 5, with 12 of 23 papers at level 3, and 10 papers name no deployment target at all. Figure 5 plots maturity against evidence strength and prints the maturity axis' own five-rater α of 0.41 on the figure, because that facet is the weakest measurement in this study and a reader should not have to take the axis on trust. The distribution is reported; no claim is built on it.

The volume has no dominant domain. The application-domain facet spreads across 7 categories with the largest holding 4 papers, and the technique facet across 10. The measurement supports no dominant theme, and §4.3 has already shown that the conventional methods for identifying one would have returned an empty graph.

Table 4
The facet matrix. **F1** primary application domain; **F2** primary technique; **F3** data modality; **F4** evaluation strategy; **F5** dataset provenance; **F6** artefact availability; **F7** deployment target; **F8** maturity, TRL-like. Codes are defined in Appendix B; the evidence score is the additive rubric over F4, F5 and F6 published there.

paper	F1	F2	F3	F4	F5	F6	F7	F8	concepts	evidence
[147]	D7	T7	M6	E2	P3	A3	G2	3	3	4
[149]	D8	T9	M8	E5	P3	A3	G2	4	4	3
[161]	D7	T5	M8	E2	P6	A3	G2	2	4	2
[162]	D7	T7	M6	E4	P6	A3	G7	3	5	1
[148]	D8	T7	M4	E4	P5	A3	G4	3	5	2
[141]	D6	T5	M5	E2	P3	A3	G8	2	3	4
[146]	D5	T7	M8	E2	P6	A3	G4	3	4	2
[142]	D6	T5	M5	E3	P4	A3	G5	2	4	2
[144]	D5	T1	M1	E1	P1	A3	G3	3	9	4
[145]	D6	T6	M3	E3	P4	A3	G8	2	6	2
[143]	D6	T8	M7	E6	P5	A3	G8	1	8	2
[152]	D2	T5	M8	E2	P4	A3	G8	2	8	3
[156]	D3	T5	M1	E4	P3	A3	G6	3	8	3
[151]	D4	T2	M1	E1	P1	A3	G3	3	11	4
[150]	D8	T9	M3	E4	P3	A3	G2	2	3	3
[157]	D3	T8	M7	E6	P5	A3	G8	1	4	2
[158]	D3	T1	M2	E1	P2	A3	G8	3	8	3
[159]	D5	T1	M1	E1	P1	A3	G1	4	9	4
[160]	D4	T1	M1	E1	P3	A1	G8	3	7	6
[153]	D2	T4	M4	E4	P6	A3	G1	3	10	1
[154]	D8	T10	M4	E2	P2	A1	G8	3	4	5
[155]	D2	T8	M7	E6	P5	A3	G8	1	5	2
[5]	D4	T3	M3	E2	P3	A3	G8	3	7	4

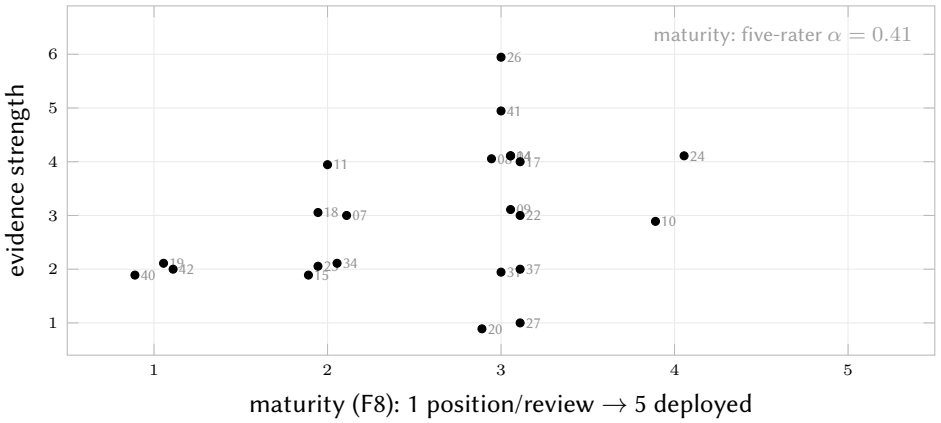


Figure 5: Maturity against evidence strength, one mark per paper, labelled by paper number and jittered deterministically so that ties are visible. Nothing in the volume reaches the top maturity level and 10 papers name no deployment target at all. The maturity axis carries its own five-rater agreement on the plot: at $\alpha = 0.41$ it is the weakest measurement in this study, so the distribution is reported and no claim is built on it.

6. The papers

Each paper is reviewed within the cluster assigned to it in §5 and is accompanied by a figure. The reviews describe what each paper does, what it measured, and what a reader should treat with care; where the discussion bore on a paper, that account draws on the coding in §7. Two papers were accepted but not presented and one was pre-recorded, so their figures come from their own sources, as their captions record.

Critical remarks are specific by intention. Every criticism made here is one the authors could act on in a subsequent version.

6.1. Language models as software substrate

Zhakhalov [152] argues that the interface language an agent uses to act, whether imperative code, RPC function calls or a declarative query, is an architectural decision with measurable consequences, and makes the case for GraphQL. The argument runs on three tracks (figure 6). On cost, a representative user–orders task shrinks from 63 request tokens to 32 and from 245 operational tokens to 175, because one typed query replaces several round trips. On reliability, schema validation occurs *before* execution and returns structured diagnostics, giving a self-correction loop that does not depend on bespoke prompt engineering. On security, a schema-bounded language acts as a native sandbox: no arbitrary-code path exists to be reached, and prompt injection is confined to a query surface already governed by depth and complexity limits.

The contribution is a reframing of the agent's "act" step as typed program synthesis against a capability algebra. It is a reframing that travels, since GraphQL is standardised and general coding models emit it without fine-tuning, and this is the strongest available form of the paper's model-agnosticism claim.

The paper has no evaluation. Its token counts come from a single worked task and not from a benchmark; there is no measured success rate against a function-calling baseline over a suite of tasks, no error taxonomy drawn from real self-correction loops, and no adversarial test of the injection claim. The security argument is sound as architecture and unquantified as engineering. Twenty tasks and a baseline would make it a considerably stronger paper, and the discussion pressed on precisely that point.

Baran et al. [153] builds what Zhakhalov [152] argues for, without citing it or knowing of it (figure 7). A conversational e-commerce assistant allows a language model to discover products by emitting Cypher against a Neo4j graph of the catalogue, with validation placed in Model Context Protocol servers so that the model never touches the database directly. Around that core the paper is carefully engineered: a change-data-capture pipeline over Debezium and Kafka keeps Neo4j consistent with the transactional PostgreSQL store, dialogue state persists across turns, and the graph encodes explicit product relationships, so that multi-hop recommendations can be explained as well as produced.

The architecture is the result and it is a good one: modular, with a clear separation between the storefront, the graph, the model and the protocol layer, and a demonstration that follows a product-discovery session end to end. The grounding argument is more than rhetorical, since an answer assembled from graph relationships can be traced back to them.

The paper lists what is missing: no concurrent-load testing, no A/B validation against the existing search interface, no penetration testing of the security mechanisms, no user study. Listing them does not fill them: the central claim that the assistant "reduces search effort" is precisely what a user study would have to establish, as the paper's own conclusion concedes. The paper also identifies prompt-injection prevention, output validation, rate limiting and query-complexity scoring as future work, which is the hardening Zhakhalov [152] treats as the point of using a schema at all. Read together, the two papers state a shared position and divide the labour between arguing it and building it. The discussion raised context-window strategy across long sessions, an unresolved engineering problem for this paradigm and a subject of §9.

Lytvynenko et al. [154] asks what preprocessing does to a sociological result, and answers with a number that should give any computational social scientist pause: of 990 Ukrainian news headlines, 107, or 10.8 %, change their compound sentiment value according to nothing but whether the pipeline lemmatised or stemmed (figure 8). The design suits the question. Two parallel corpora are each aligned with a correspondingly processed sentiment dictionary, so that the comparison isolates morphological normalisation instead of confounding it with dictionary mismatch, and the choice of Ukrainian is deliberate, since a morphologically rich language is where the two strategies diverge most.

The framing is the paper's strength: preprocessing is treated as a substantive methodological decision and not a technical preliminary, and NLP as a bridge between interpretive and quantitative sociology.

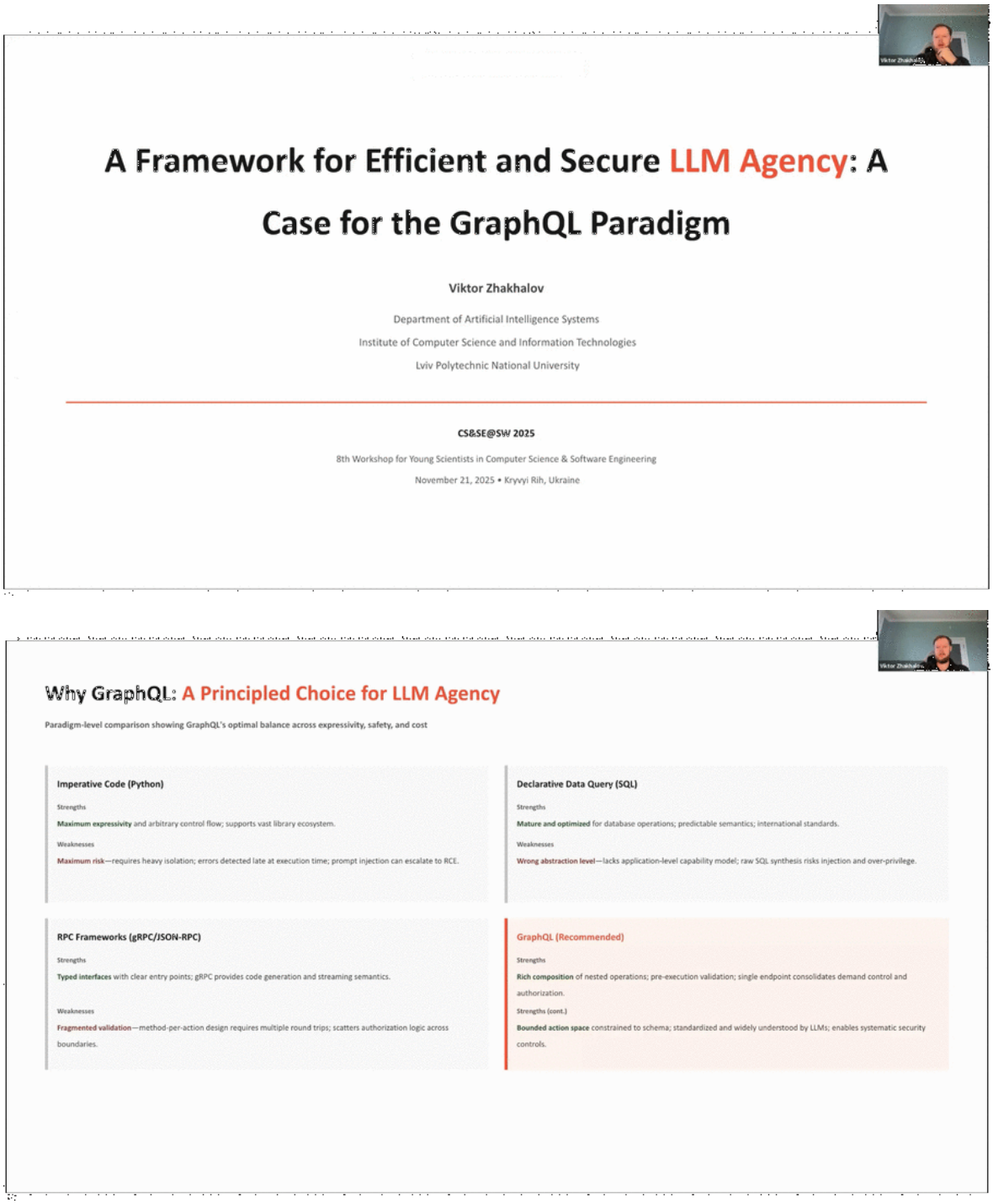


Figure 6: Excerpts from the presentation of Zhakhalov [152] (panels A–B; discussion 01:40:43–01:49:31).

The workflow is documented well enough to repeat, and the paper states plainly that its custom dictionary is a demonstration and not a validated instrument.

Three limits matter. The sentiment lexicon is the authors' own and unvalidated, so the 10.8 % measures divergence under *this* lexicon and not a general rate. Headlines are a short and atypical genre, and the OCR step adds noise that the paper acknowledges without quantifying. Lexicon-based scoring is also a weak baseline in 2025: the obvious comparison, a transformer classifier fine-tuned for Ukrainian, is named as future work but not run, and it would have shown whether preprocessing sensitivity is a property of lexicon methods specifically or of the task. The paper was accepted but not presented, so

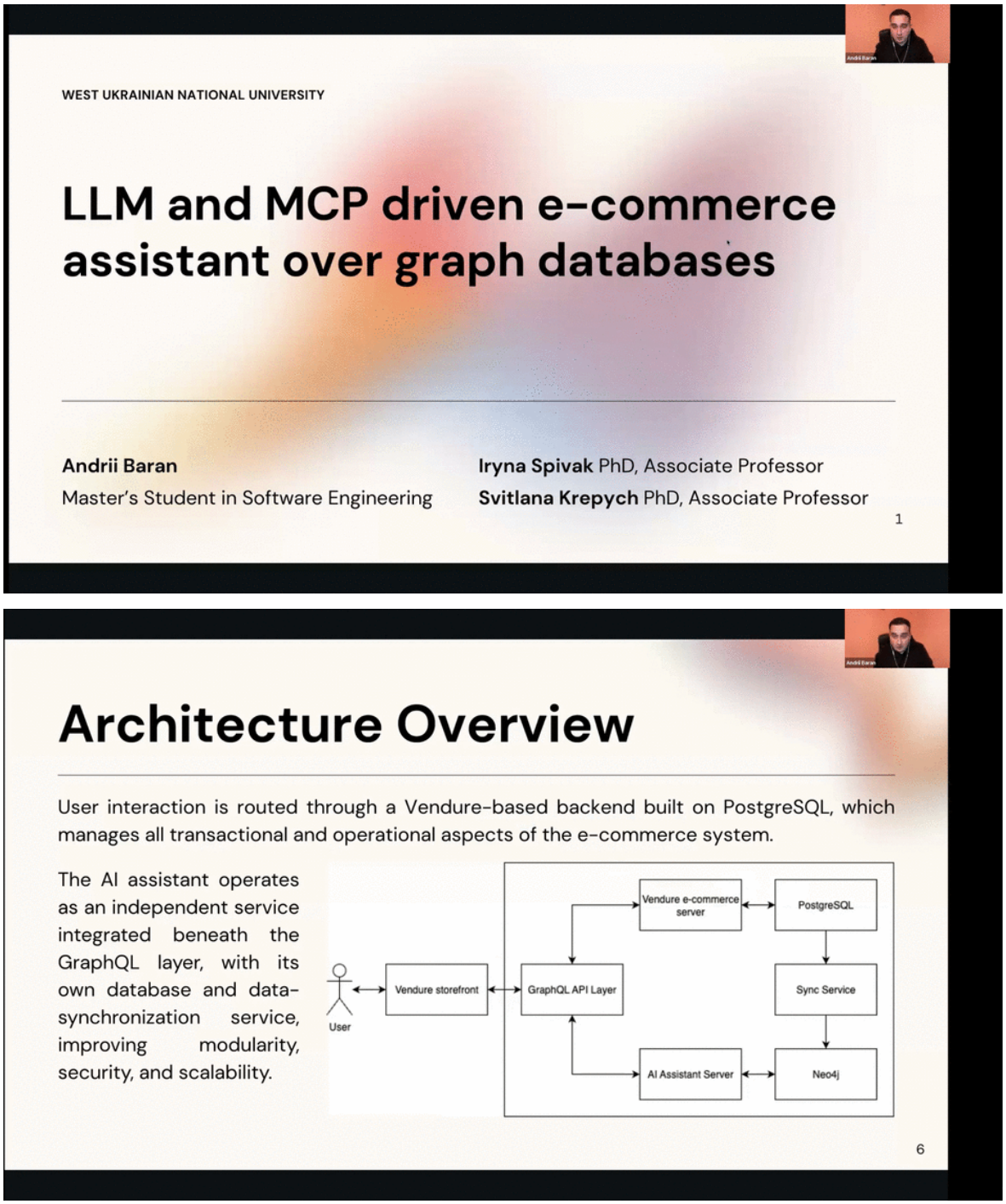


Figure 7: Excerpts from the presentation of Baran et al. [153] (panels A–B; discussion 03 : 36 : 14–03 : 47 : 23).

it has no discussion and its figures are its own. It is one of only 2 papers in the volume to release an artefact.

Slobodianiuk and Semerikov [155] is a systematic review of deep-learning text generation over 2022–2024, and it does the standard job competently: 43 Scopus articles, a stated search strategy, and a finding that Transformer variants dominate while attention mechanisms and controllable generation gain ground (figure 9). Its more interesting feature lies elsewhere.

The search was executed in February 2024. Instead of publishing a 2026 paper whose evidence stops two years earlier and leaving the reader to discover the fact, the paper adds an explicitly declared *non-systematic* narrative section covering 2024–2026 and reports what it finds: Transformer dominance

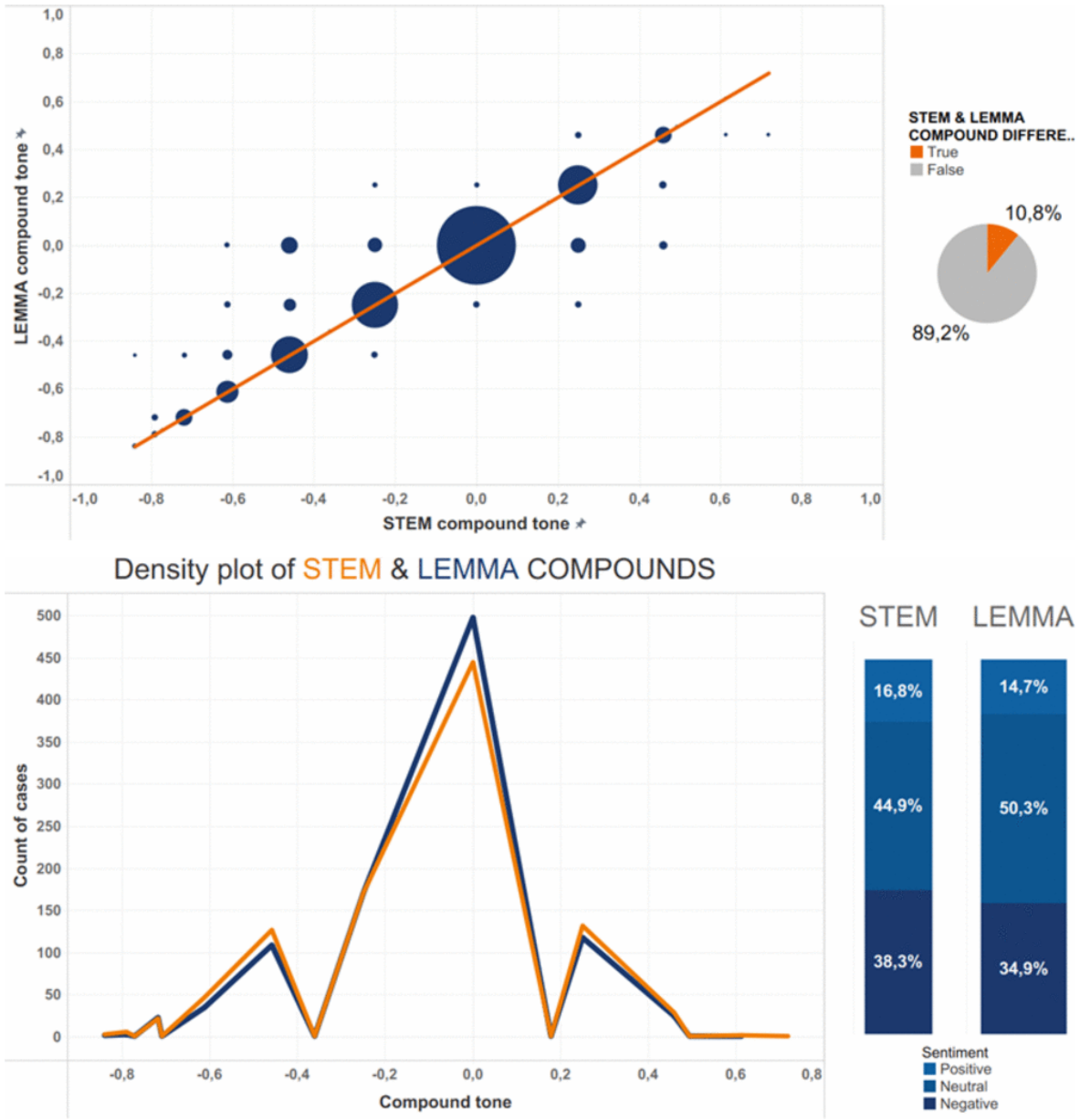


Figure 8: Figures from Lytvynenko et al. [154] (panels A–B).

persisted, but the locus of innovation moved off the architecture and onto post-training, inference-time computation, retrieval and evaluation. The conclusion then instructs a 2026 reader to apply the systematic findings with that qualification, the decisive variables no longer being the choice of backbone.

The construction deserves imitation. A declared boundary between what was retrieved systematically and what was added narratively is better practice than either a silent update or a stale conclusion. The cost is that the added section cannot support the same claims as the reviewed corpus, which the paper does not attempt.

The weaknesses are the ordinary ones for the genre. A single database, Scopus, was searched with no supplementary source, which for this field's arXiv-first literature is a real coverage limit; 43 articles is a small yield for a three-year window and suggests tight inclusion criteria whose effect is not tested; and the included studies are assessed for content but not for quality. The finding about where innovation moved nonetheless supports the argument of §9, and it arrives from a different direction than the two agent papers do. The paper was accepted but not presented.

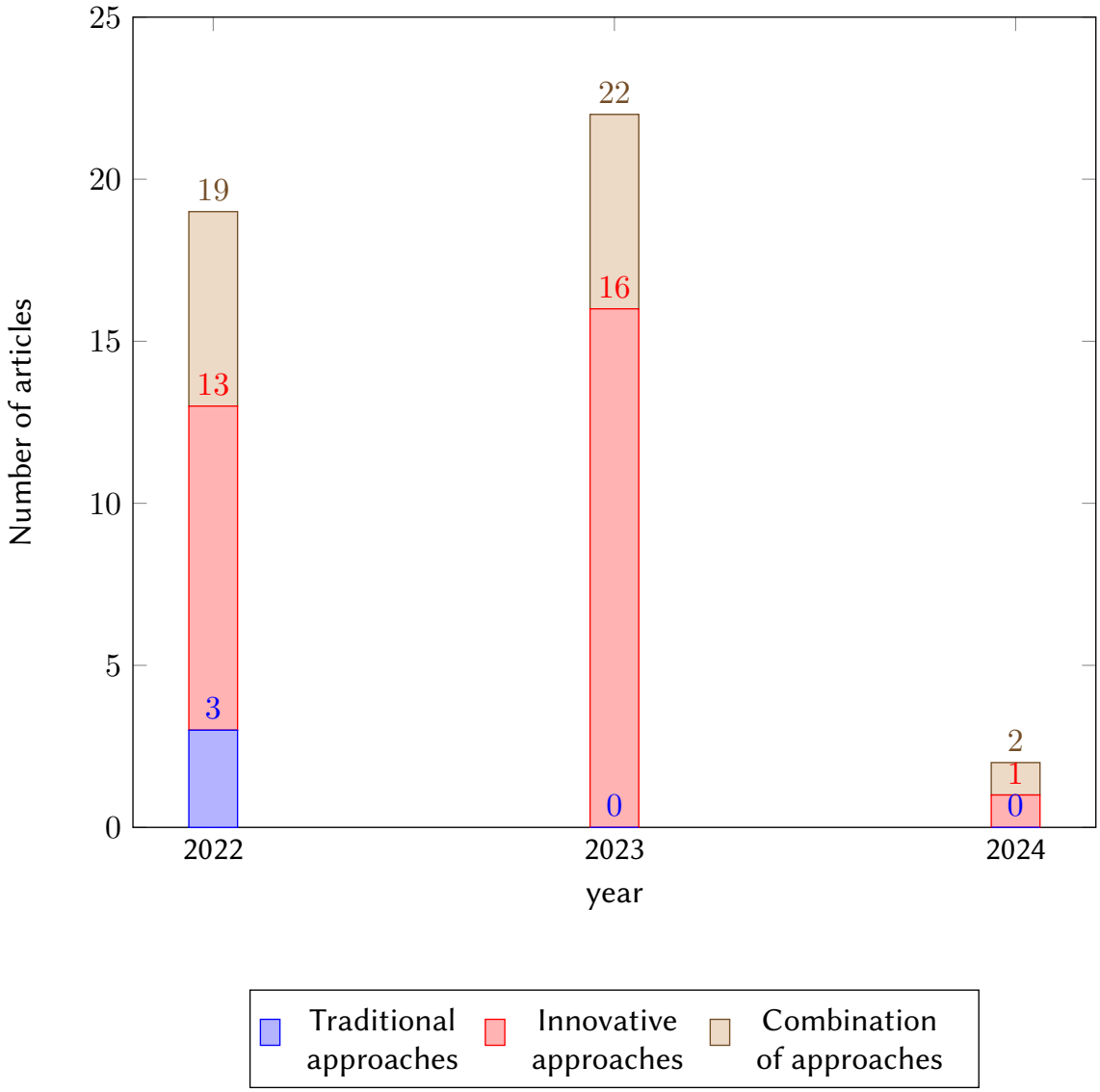


Figure 9: Figure from Slobodianiuk and Semerikov [155] (panel A).

6.2. Airborne sensing and geospatial intelligence

Rolik et al. [156] treats demining survey as a systems problem, arguing that a UAV carrying infrared sensors is *the only safe way* to characterise a newly liberated agricultural area before people walk into it (figure 10). The framing earns the paper its place in the volume, and the engineering is worked through to numbers: for a demining brigade's configuration, an object diameter of 0.12 m, a minimum slope angle of 150°, smallest viewing angles of 80° and 40°, 720 pixels across the narrow field, five pixels needed to declare a detection, and a 150 mm lens. The flight geometry is derived from these rather than asserted.

The paper's most creditable feature is what it reports about its own errors. Camera calibration is performed and not assumed; spatial points are projected into the image plane with OpenCV; error is shown to grow with distance from the optical axis, so that peripheral detections require further correction; and automatic exposure is shown to cost contrast as the aircraft descends, which the processing then accounts for.

The field trial is missing. What is presented is a structural and functional model with calibration and geometry, not a deployment with detection rates on real ordnance, and there is no false-positive

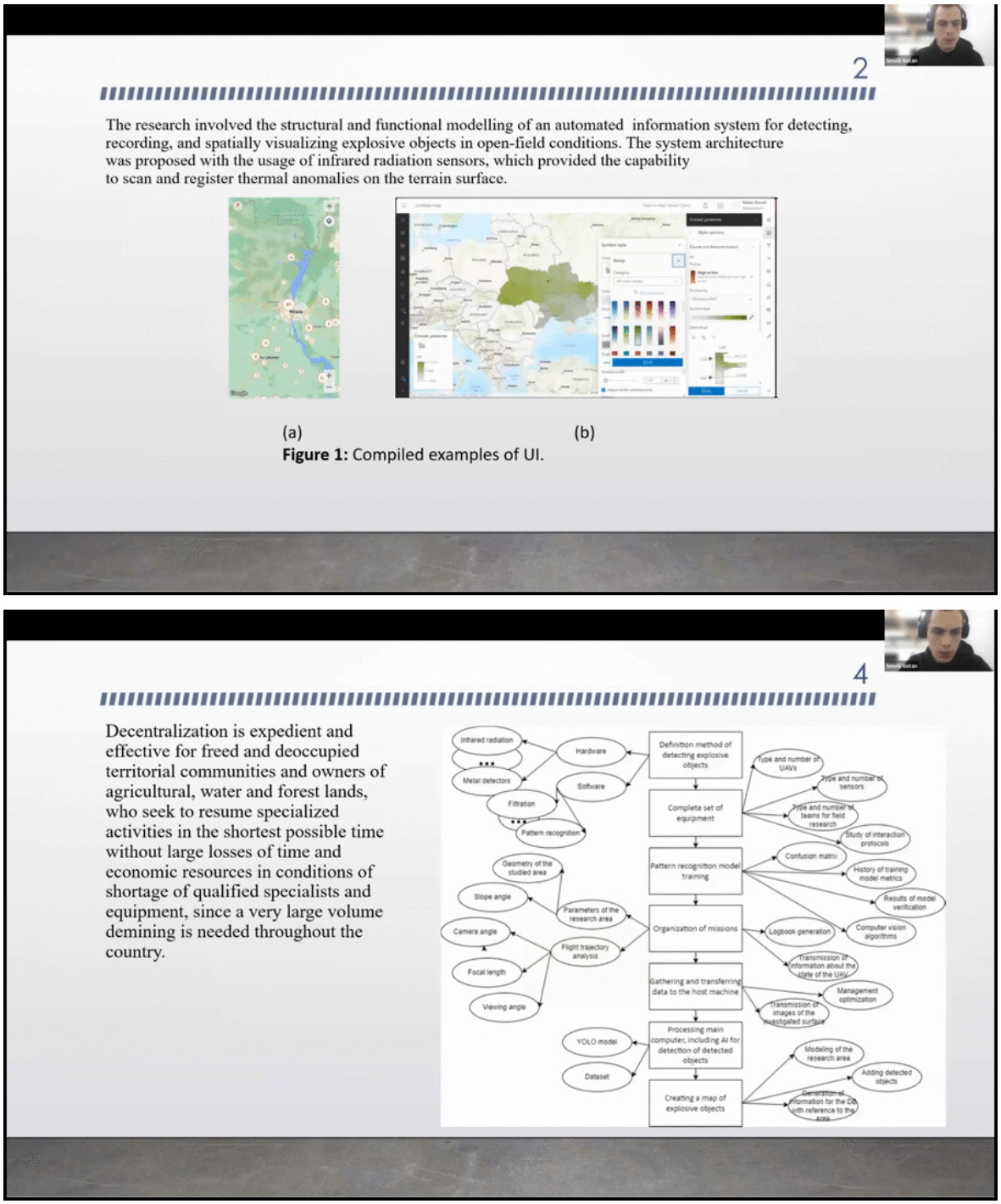
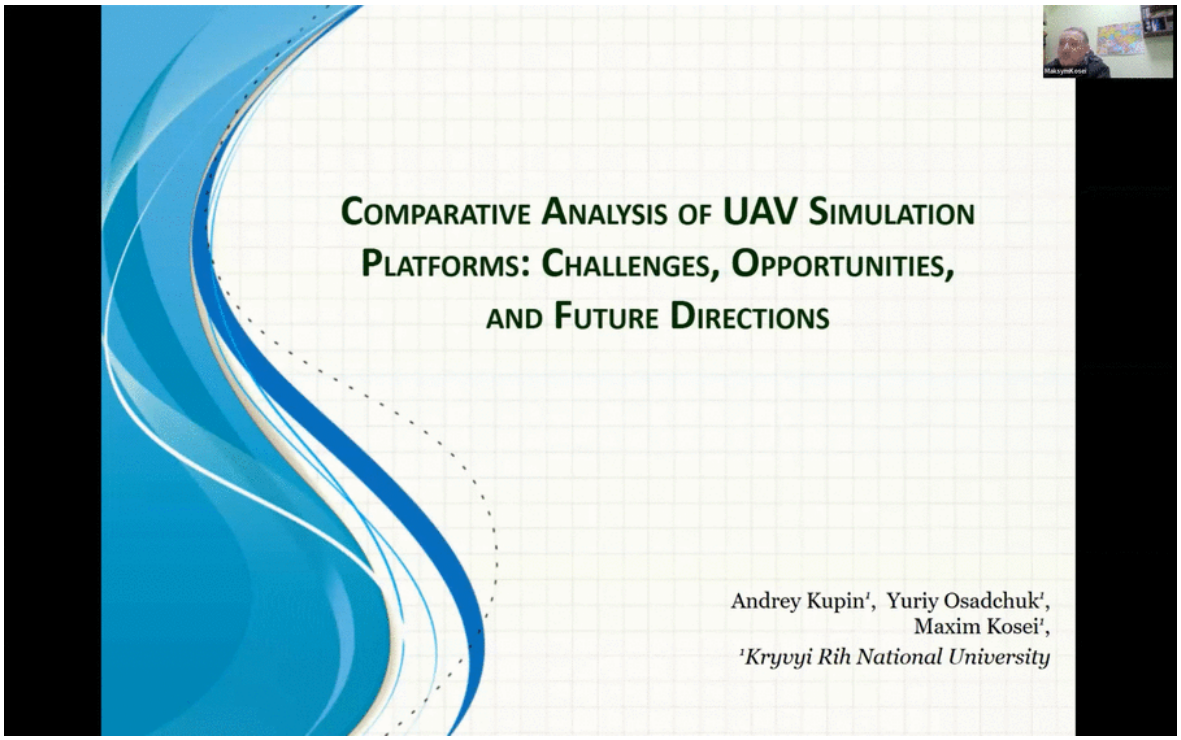


Figure 10: Excerpts from the presentation of Rolik et al. [156] (panels A–B; discussion 01:55:40–02:04:11).

analysis. In this application that is the number that matters most, since a missed object and a spurious one carry entirely different costs. There is also no comparison with the survey methods currently in use. The discussion pressed on the sensing assumptions.

Kupin et al. [157] is a meta-review of nine studies of UAV simulation published between 2018 and 2025, six of them already comparative surveys, and it is the volume's clearest demonstration that a secondary study can produce a finding and not merely a summary (figure 11). The finding is that the literature comparing UAV simulators cannot be read together. The surveys converge on a recurring vocabulary of evaluation axes, among them sensor support, physics fidelity, scalability, usability and maintenance, but they name and weight those axes differently, so that a platform's standing depends



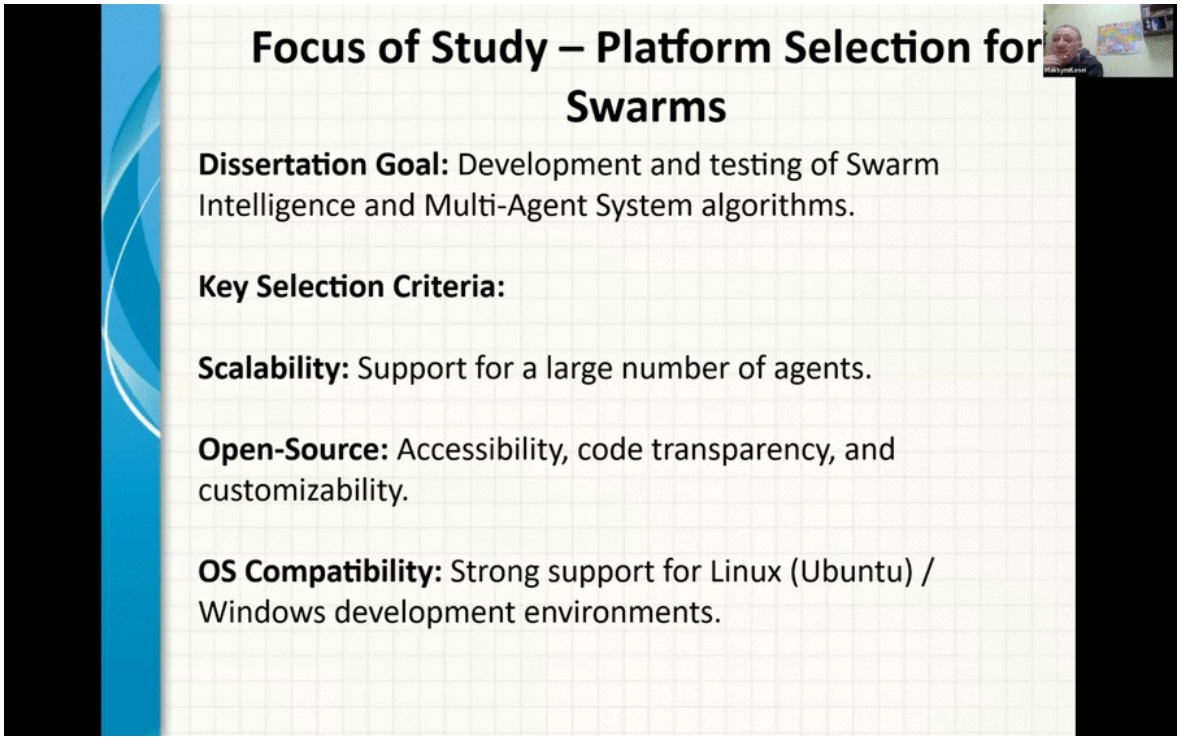


Figure 11: Excerpts from the presentation of Kupin et al. [157] (panels A–B; discussion 03 : 59 : 45–04 : 05 : 46).

on which survey one consults. Coverage is also grossly uneven, Gazebo being assessed in five of the nine studies and most platforms in one, and neither the classification schemes nor the scoring scales translate between studies.

The paper draws the appropriate conclusion: the field's gap lies not in simulator capability but in the absence of a shared benchmarking protocol. It specifies what such a protocol would have to fix, namely reference scenarios for single-vehicle navigation, swarm coordination and sensor degradation; reported metrics including wall-clock cost per simulated second and *the measured sim-to-real discrepancy*,

which no reviewed study reports reproducibly; and a declared weighting, so that a score states its own assumptions.

One incidental observation does more work than its length suggests: of seven platforms compared in one survey, the newest documentation for three dates from 2012, 2015 and 2016. A comparison that treats an unmaintained simulator as a live option is comparing feature lists rather than tools, and the paper is right that maintenance activity belongs among the evaluation axes.

The limits are inherent to the design. Nine studies is a small base, the selection criteria are not stated as a search protocol, which is a notable omission in a paper arguing for protocols, and no inter-rater check is reported on the reconstruction of each study's scope and criteria. The proposed protocol is specified but not piloted; running it on even two platforms would have converted the argument into evidence.

Chaskovskyy et al. [158] monitors forest disturbance in the Skole Beskids national park by combining Sentinel-1 SAR, Sentinel-2 multispectral imagery and a global DEM, and compares a random forest against a UNet with a ResNet-34 backbone (figure 12). The deep model wins where it should, on boundary delineation and noise suppression, at IoU 0.84, F_1 0.91 and pixel accuracy 0.96 on the validation set, and the study reports roughly 1.247 km^2 of forest change over 2023–2024, concentrated near commercial logging areas and within military forest districts.

Two design choices lift this above a routine classification study. The sensor combination is motivated, not assembled: SAR carries structural change through cloud and canopy where optical data cannot, which in a mountain park is the difference between a usable time series and a seasonal one. And the pipeline ends in an operational Web GIS, with processed data moving from Google Earth Engine to an online interface by API, so that a forest manager can see a disturbance early enough to act. The paper's claim to support earlier detection of illegal logging rests on that integration and not on the metrics.

The metrics themselves need reading with care, and the paper supplies the means to do it: an overall accuracy of 83.81 % and a Kappa of 0.676 sit well below the pixel accuracy of 0.96, which is what happens when one class dominates the scene. Reporting all of them together is good practice, though the paper could have stated more clearly which figure is the headline. Beyond that, validation is against interpreted imagery rather than field plots, a single park over a single year cannot establish transferability, and the disturbance *agent*, whether storm, harvest or illegal cut, is not identified, which is the distinction a governance user most needs. The paper acknowledges this and lists the sensors that would help.

Krak et al. [159] assembles a real-time counter-UAV pipeline from three components: YOLOv12 for detection, BYTETrack for identity maintenance, and a position-coded Transformer for short-horizon trajectory forecasting. It is evaluated on the visible-spectrum subset of the Anti-UAV benchmark, 593,802 frames of real surveillance video (figure 13). The numbers are specific and, unusually for this volume, include throughput on both accelerators and commodity hardware: mAP_{50} 91.04 %, $mAP_{50–95}$ 62.76 %, MOTA 0.79, IDF1 0.64, forecast RMSE 12.66 pixels, 55 FPS on GPU and 13 on CPU, with the tracker alone at 332 FPS on CPU.

The forecasting comparison is the paper's strongest evidence: against IMM and Kalman baselines the Transformer reduces error by roughly an order of magnitude. That is a comparison against the methods actually deployed in this application, not against a weaker neural alternative, and it is the right baseline to beat. Translating the pixel error into 0.25–0.32 m of physical scene is also the correct instinct, giving the reader a quantity they can reason about operationally.

The limitations are stated clearly and bound the result tightly. Forecasting is performed in two-dimensional image coordinates from a fixed camera, so a moving platform, which is the deployment that matters for a mobile counter-UAV system, would require different work. The forecast horizon is short. The visible-spectrum subset excludes the low-light conditions in which such a system would be most needed, and an IDF1 of 0.64 indicates that identity is not held reliably through occlusion, which a forecaster then inherits as corrupted history. The paper names most of these constraints itself.

6.3. Applied machine learning for health, environment and industry

Fedorchenko et al. [5] builds individual diets by comparing four optimisation algorithms (greedy,

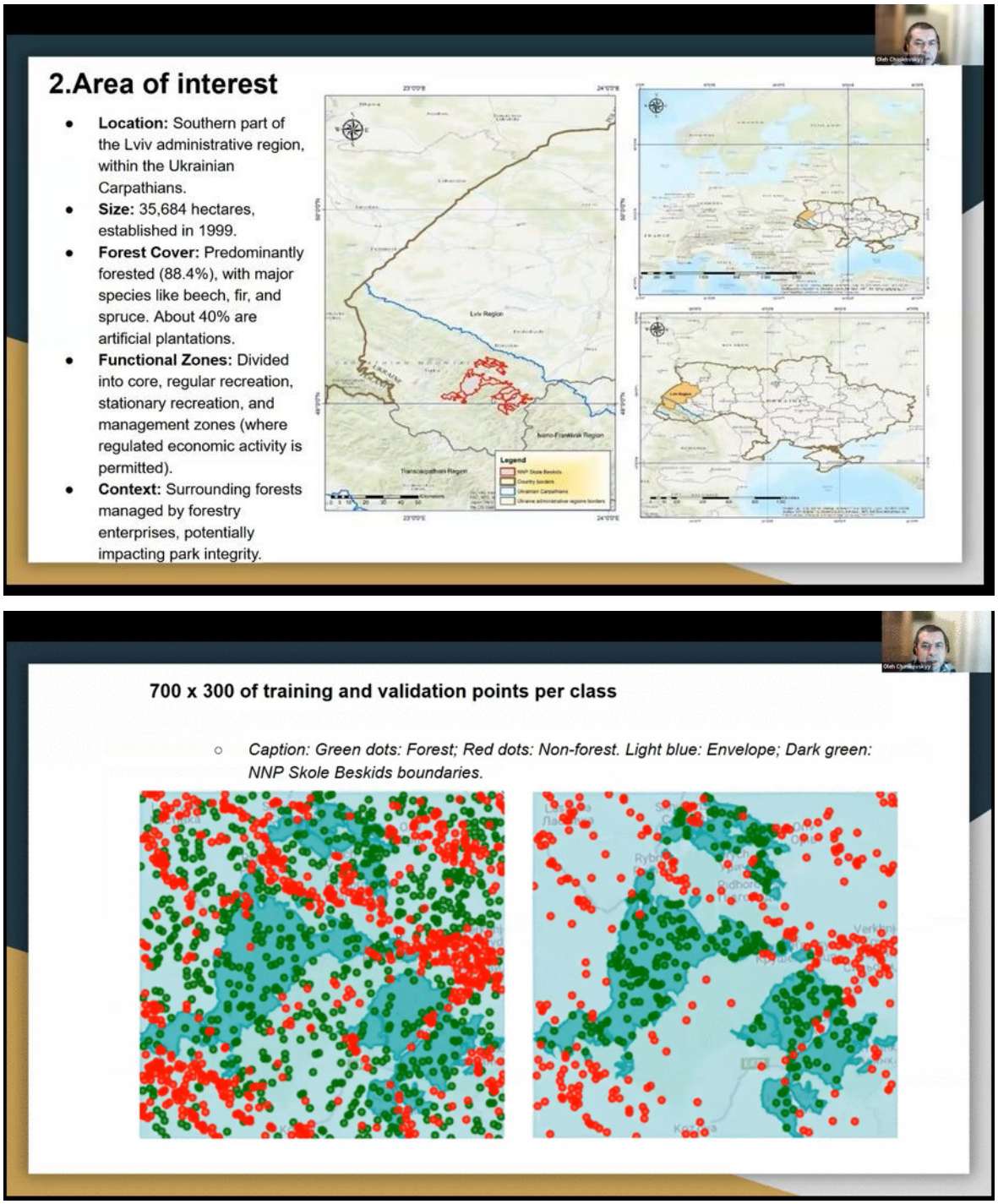
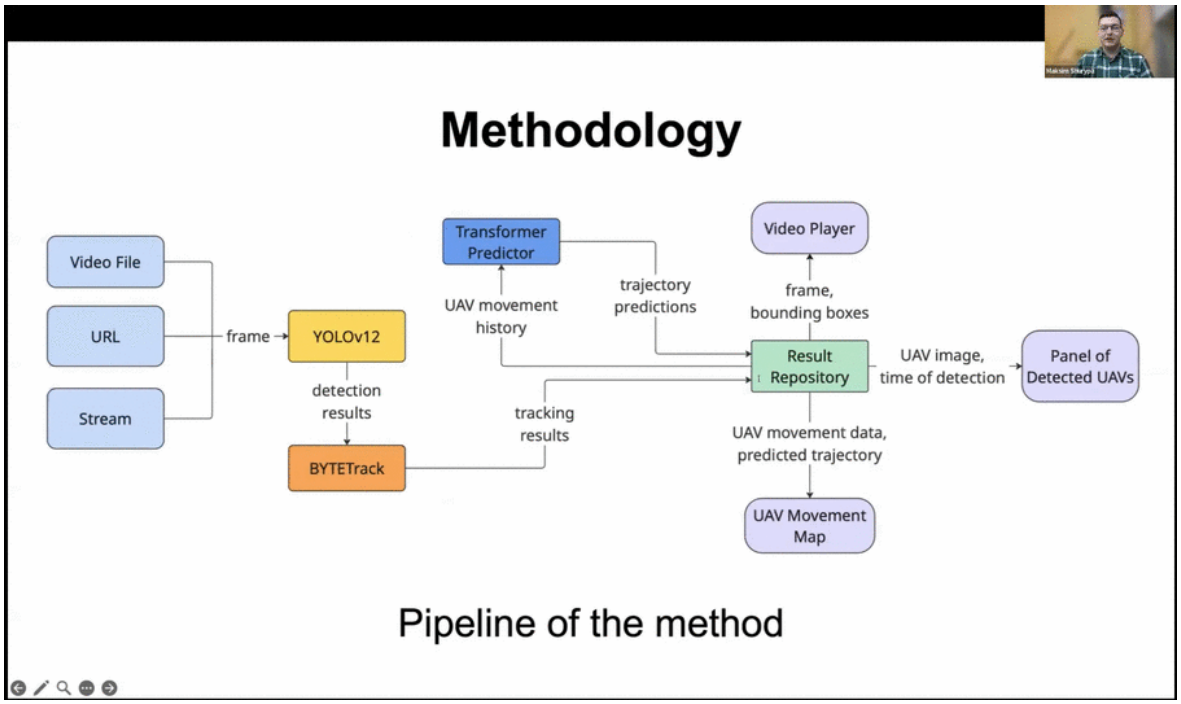


Figure 12: Excerpts from the presentation of Chaskovskyy et al. [158] (panels A–B; discussion 04:43:35–04:49:57).

linear programming, Monte Carlo and a genetic algorithm) against three predictors (random forest, XGBoost and a dense neural network) for distributing food components across a person's daily calorie requirement, derived from basal metabolism and BMI. The comparison is scored on three criteria, and the third is the interesting one: alongside accuracy and computation speed the paper counts the number of distinct acceptable diets an algorithm can produce (figure 14). For a recommender that a person has to live with, variety is a genuine quality dimension instead of a curiosity, and most papers in this space do not measure it.

The conclusion, that a hybrid of optimisation and learned prediction beats either alone, is plausible



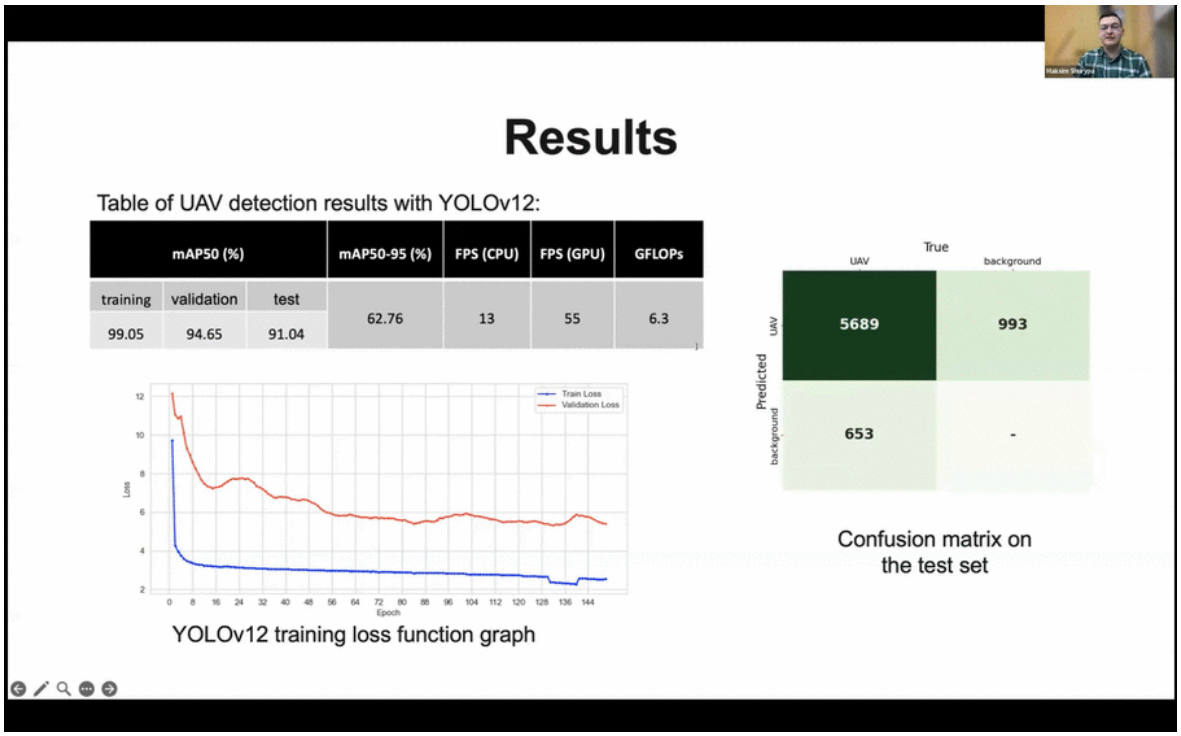


Figure 13: Excerpts from the presentation of Krak et al. [159] (panels A–B; discussion 04:59:38–05:03:30).

and is the standard finding in this literature, which is both a reassurance and a limit on the contribution's novelty. The paper's structure, seven algorithms compared on stated criteria rather than one method advocated, is the right shape for a workshop contribution.

Two things a reader should hold in mind. There is no clinical or nutritional validation: the optimiser satisfies a calorie and macronutrient specification, and whether the resulting diets are nutritionally sound, palatable or adhered to is outside what is tested. And the evaluation is on synthetic requirement profiles, not real users, so "accuracy" means fidelity to the specification, not health outcome. This is the paper presented at the workshop but published in a journal and not in this volume, and it is cited to

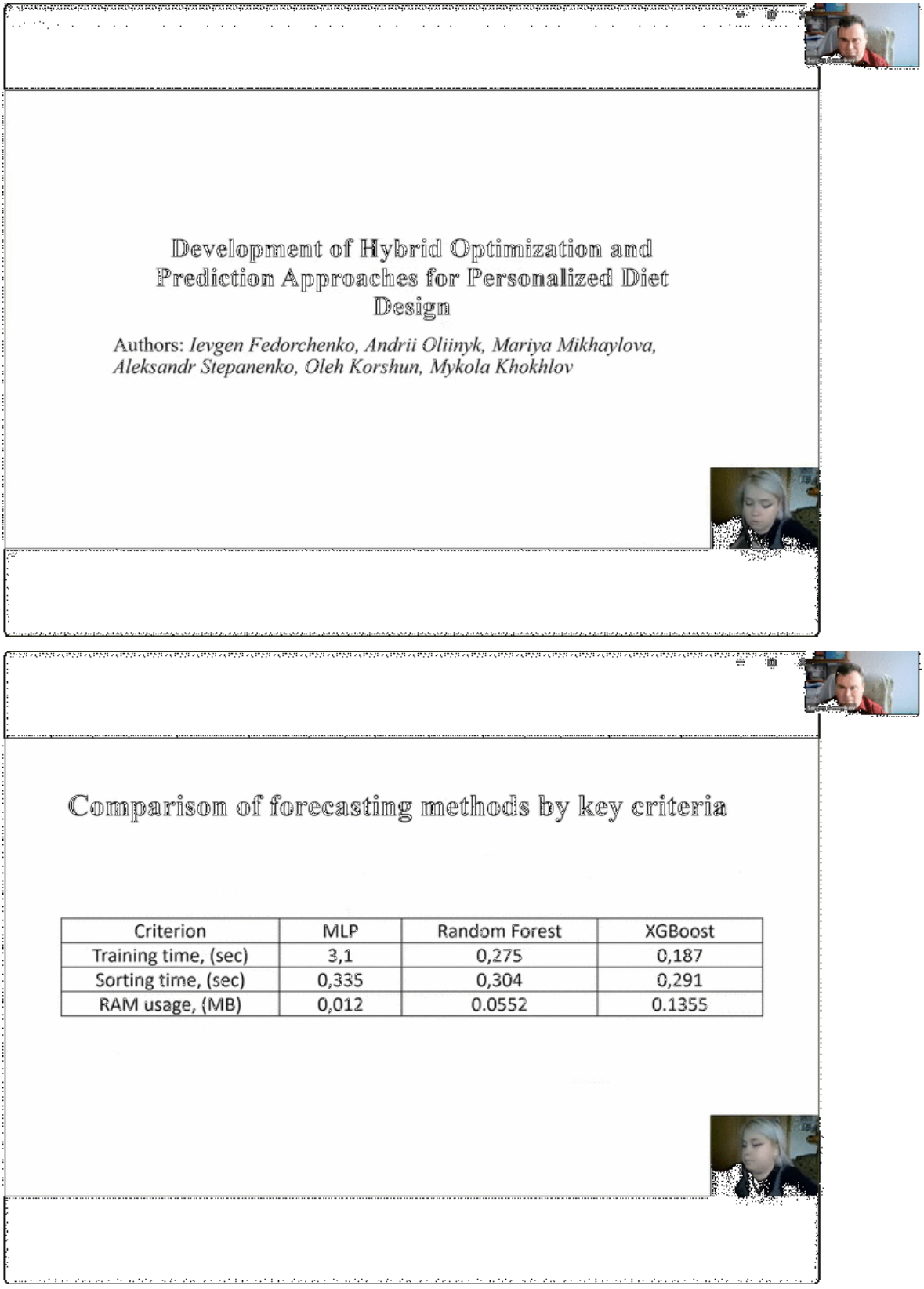


Figure 14: Excerpts from the presentation of Fedorchenko et al. [5] (panels A–B; discussion 00:36:54–00:41:05).

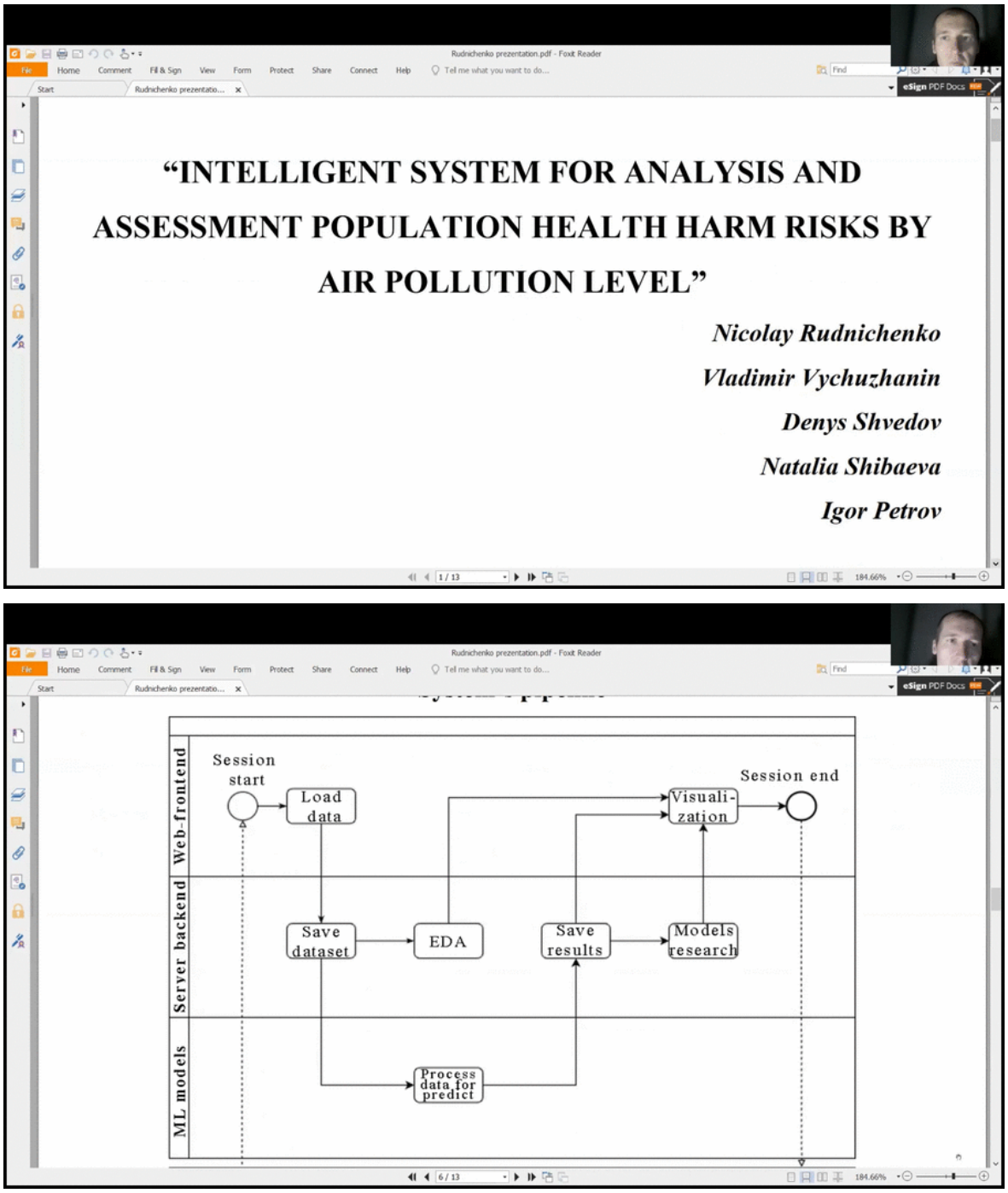


Figure 15: Excerpts from the presentation of Rudnichenko et al. [151] (panels A–B; discussion 03:04:43–03:09:00).

that published record; note that its published title differs from the one on its workshop source.

Rudnichenko et al. [151] classifies air-quality observations into six AQI bands as a proxy for population health risk, and its methodological discipline is what distinguishes it (figure 15). Six models (decision tree, SVM, random forest, XGBoost, CNN and LSTM) are trained on *identically* preprocessed data and compared at the evaluation stage rather than chained into a pipeline, which is the design that actually licenses a statement about which model is better. Random forest and CNN lead at 0.988 and 0.987 accuracy with macro-F_1 of 0.987 and 0.986, the LSTM follows at 0.980, and the single decision tree falls to 0.856, and the paper explains why, and not just reporting it: the tree fails to separate the two most severe classes, which are the ones a health authority cares about.

The paper then makes the argument that decides its own design. Training time runs from 3 s for the tree to 67 s for the LSTM, so the random forest is the better engineering choice, not merely the joint-most accurate model. Reaching for the cheapest model that is indistinguishable at the top is the correct reasoning and is rarer than it should be. The system is delivered as a working client–server application with a Flask backend and React dashboard offering exploratory analysis, inference and model comparison.

The caution is about the data instead of the models. Accuracies of 0.988 on six-class environmental data invite suspicion of leakage or of a train/test split that shares near-duplicate observations, and the paper does not report the split protocol in enough detail to dismiss that. The primary dataset is an image dataset from India and Nepal cross-checked against structured measurements, so transfer to a Ukrainian city, the implied deployment, is untested. And AQI class is a proxy for harm, not harm; no health outcome is measured. One ask in discussion.

Markevych et al. [150] studies how motivational factors affect the performance of staff working in inclusive educational settings, and builds an expert system over questionnaire data to identify which needs matter for which individual (figure 16). The framing is sound and the problem is real: inclusive-setting work demands psychological resilience and internal readiness on top of qualification, and a management instrument that surfaces individual need, and not merely an aggregate, has a clear use.

This is the volume's clearest example of a paper whose method sits uneasily with its ambition. The instrument is a questionnaire; the analysis is expert-system rules over its responses; and the outcome variable, staff effectiveness, is not independently measured, so what the system relates is self-reported motivation to self-reported or expert-judged performance. The paper does not report the questionnaire's construction, its reliability, its sample size or its respondent population in the detail needed to assess any of this, and no validation against an external performance measure is offered. The claim that the approach improves educational quality is therefore a hypothesis the paper motivates, not a result it establishes.

Read as an instrument-design paper it is more defensible than read as an empirical one, and the useful next step is small and obvious: report the psychometrics. The cartography pairs this paper with Bodnenko et al. [149] across cluster boundaries because both are questionnaire-and-expert-method studies of practitioners (§5.1), a similarity the domain-based partition hides and the method-based one finds. Its discussion was the volume's most critical: four asks, three of them methodological challenges, plus a camera-ready directive.

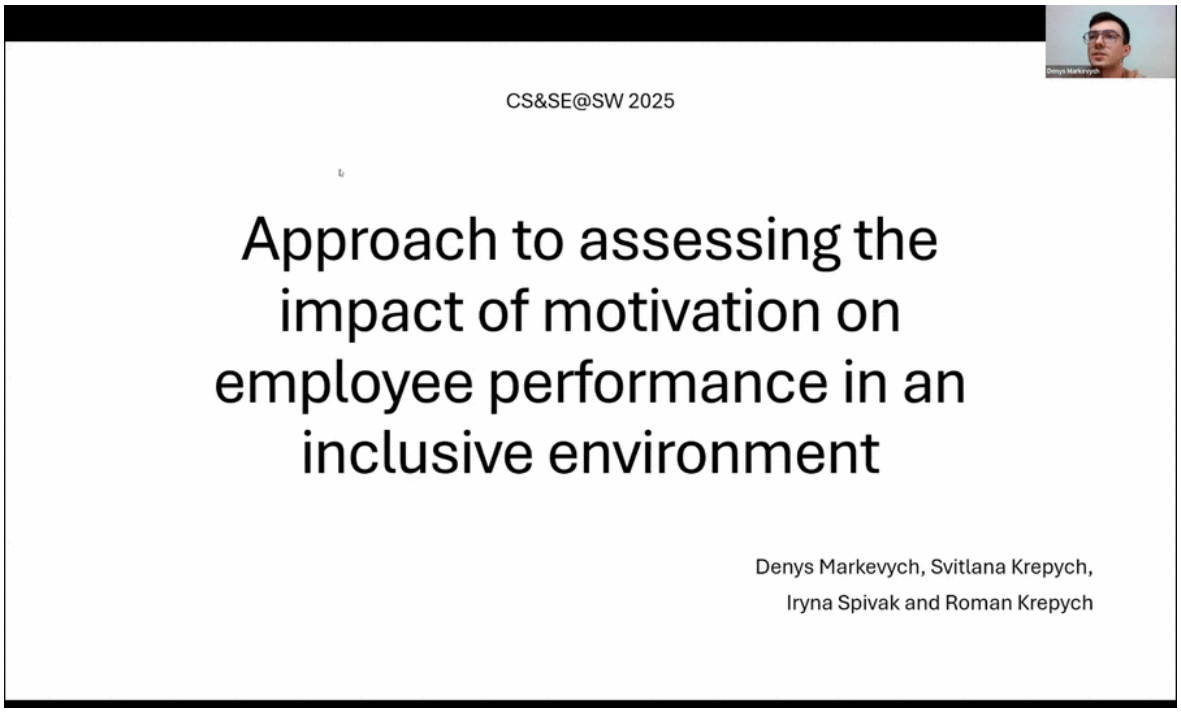
Mazurets et al. [160] determines the synthetic-to-natural fibre ratio of a fabric from photographs of its macrostructure, positioning the method explicitly as an alternative to expensive spectral analysis (figure 17). Three architectures are compared (ViT-B/16, EfficientNet-B0 and ConvNeXt-Tiny) and the transformer wins at 98.2 % accuracy and 98.5 % F_1, above published CNN and spectral baselines. Classification is into three bands (30–50 %, 50–70 %, above 70 %), which the paper justifies by use and not by convenience: those are the intervals that matter for sorting, labelling compliance and recyclability assessment.

Choosing the granularity a downstream user needs, instead of the finest the model can support, is a mature design decision and the paper is right to foreground the applications. It is also one of only 2 papers in the volume to release an artefact, publishing its fabric dataset, which given §5.8 is worth emphasising.

The evidence is thinner than the accuracy figure suggests. Three coarse bands are an easier problem than a continuous estimate, and 98.2 % on three classes from one dataset says more about the dataset's separability than about the method's robustness: the paper itself names worn materials, heterogeneous structures and variable operating conditions as untested, which is to say the conditions of the sorting line it targets. There is no controlled comparison against an actual spectral instrument on the same samples, which is the comparison the cost argument needs.

6.4. Signals, security and resource-constrained computing

Quan et al. [141] improves a generalized sidelobe canceller for microphone-array speech enhancement



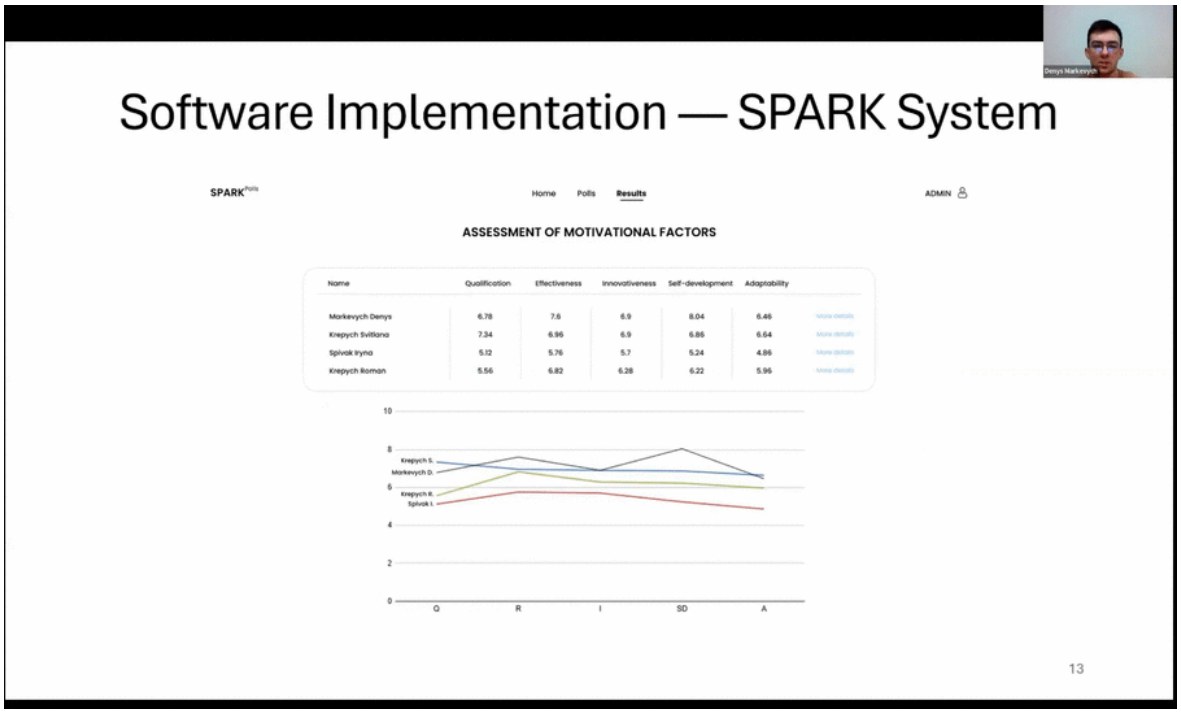


Figure 16: Excerpts from the presentation of Markevych et al. [150] (panels A–B; discussion 03:15:47–03:22:30).

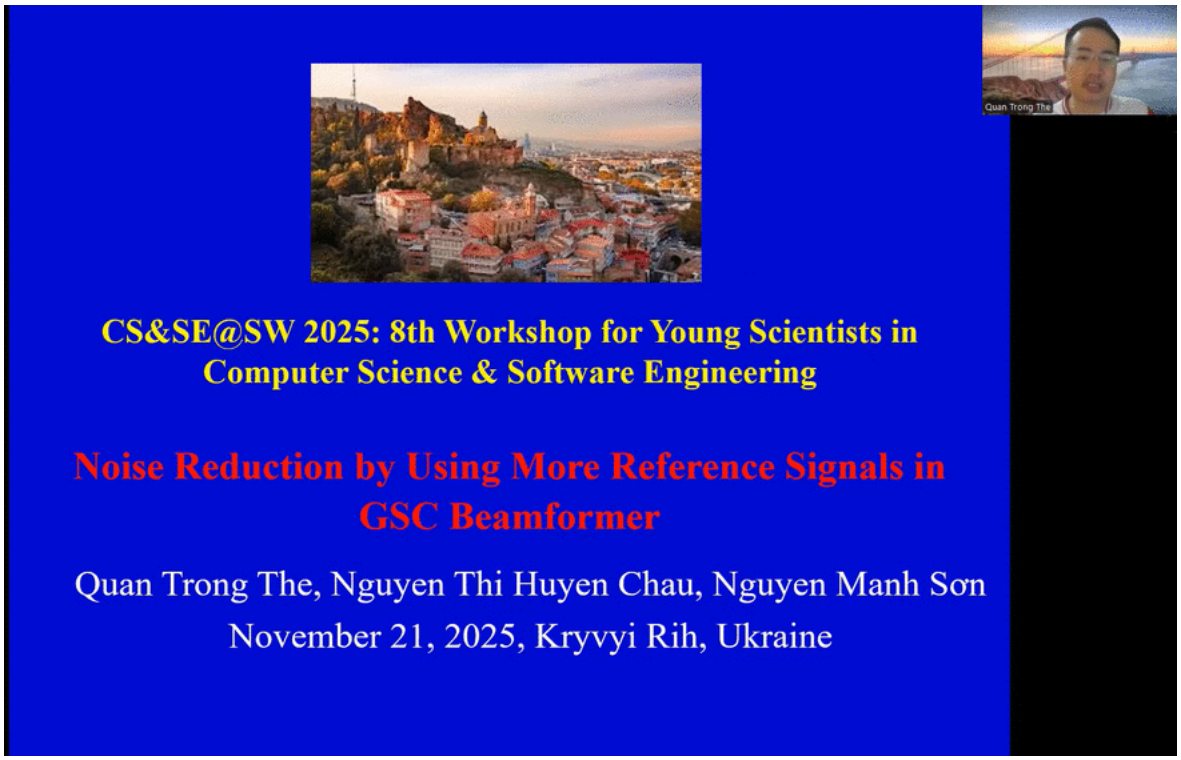
by modifying the steering vector to obtain more reference signals, targeting exactly the conditions under which the classical formulation degrades: a talker who moves their head, microphones with mismatched sensitivities, sampling-rate error and array displacement (figure 18). The reported gain is an SNR improvement from 7.4 to 8.8 dB, with a reduction in the musical noise that makes beamformer output unpleasant to listen to.

The paper is admirably focused on the realistic-conditions problem rather than on the ideal case, and the diagnosis is the right one: robustness failures in GSC beamforming are usually steering mismatch, not insufficient adaptation. Framing the fix as a preferred steering vector keeps it implementable inside

Figure 17: Excerpts from the presentation of Mazurets et al. [160] (panels A–B; discussion 05:11:40–05:15:12).

existing multi-channel systems.

The evidence, however, is one aggregate number. A 1.4 dB improvement is reported without a confidence interval, without the number of recording conditions it averages over, and without the perceptual or intelligibility measures (PESQ, STOI) that the paper's own motivation invokes when it talks about speech quality and intelligibility. Nor is there a comparison against the standard robust alternatives, such as diagonal loading or a robust MVDR formulation, which is what a reader needs in order to place the contribution. The idea is sound and the reporting is under-powered relative to it. This paper was shown but not discussed, its authors were absent at the start of the session and the



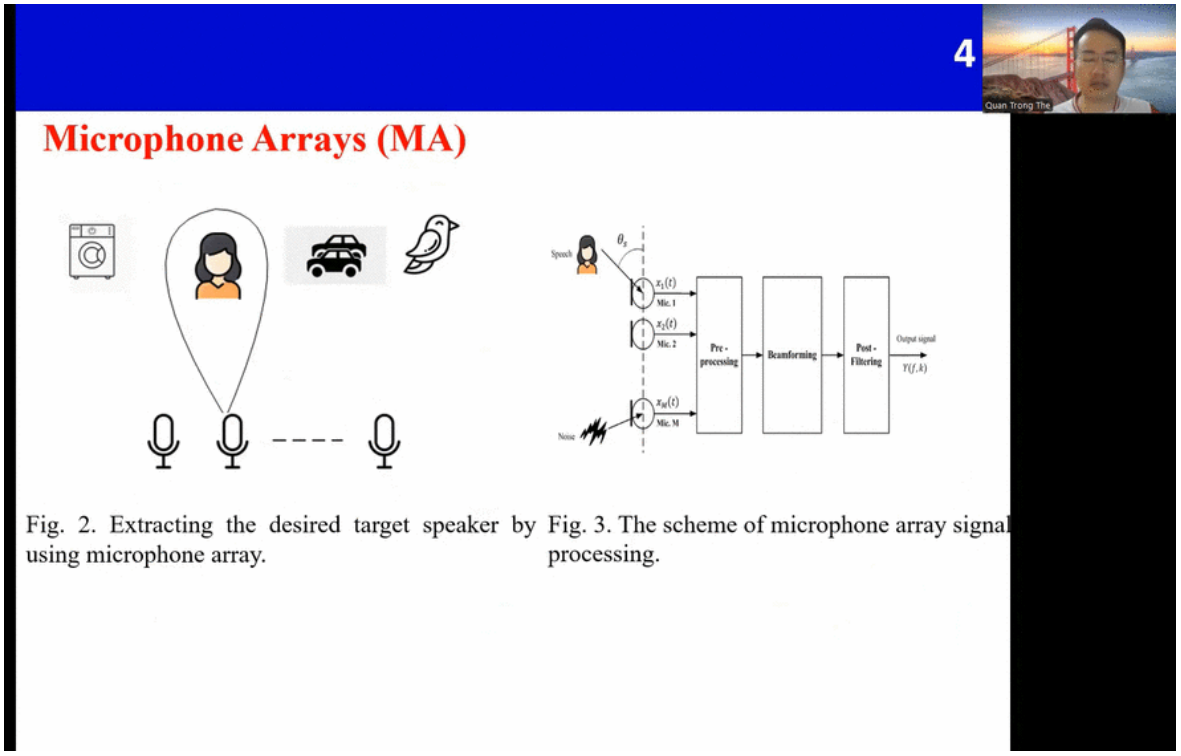


Fig. 2. Extracting the desired target speaker by using microphone array.

Fig. 3. The scheme of microphone array signal processing.

Figure 18: Excerpts from the presentation of Quan et al. [141] (panels A–B; shown as slides only: the authors could not attend their slot).

slides were presented at the end, so it has no Q&A, and its figure is drawn from slide-only captures.

Hrytsai et al. [144] develops a face-authentication module and does the comparative groundwork properly: OpenCV, Dlib, face_recognition, DeepFace and InsightFace are compared on accuracy, performance and implementation complexity before one is chosen, and the survey of knowledge-based, possession-based and biometric authentication situates the choice (figure 19). The implementation

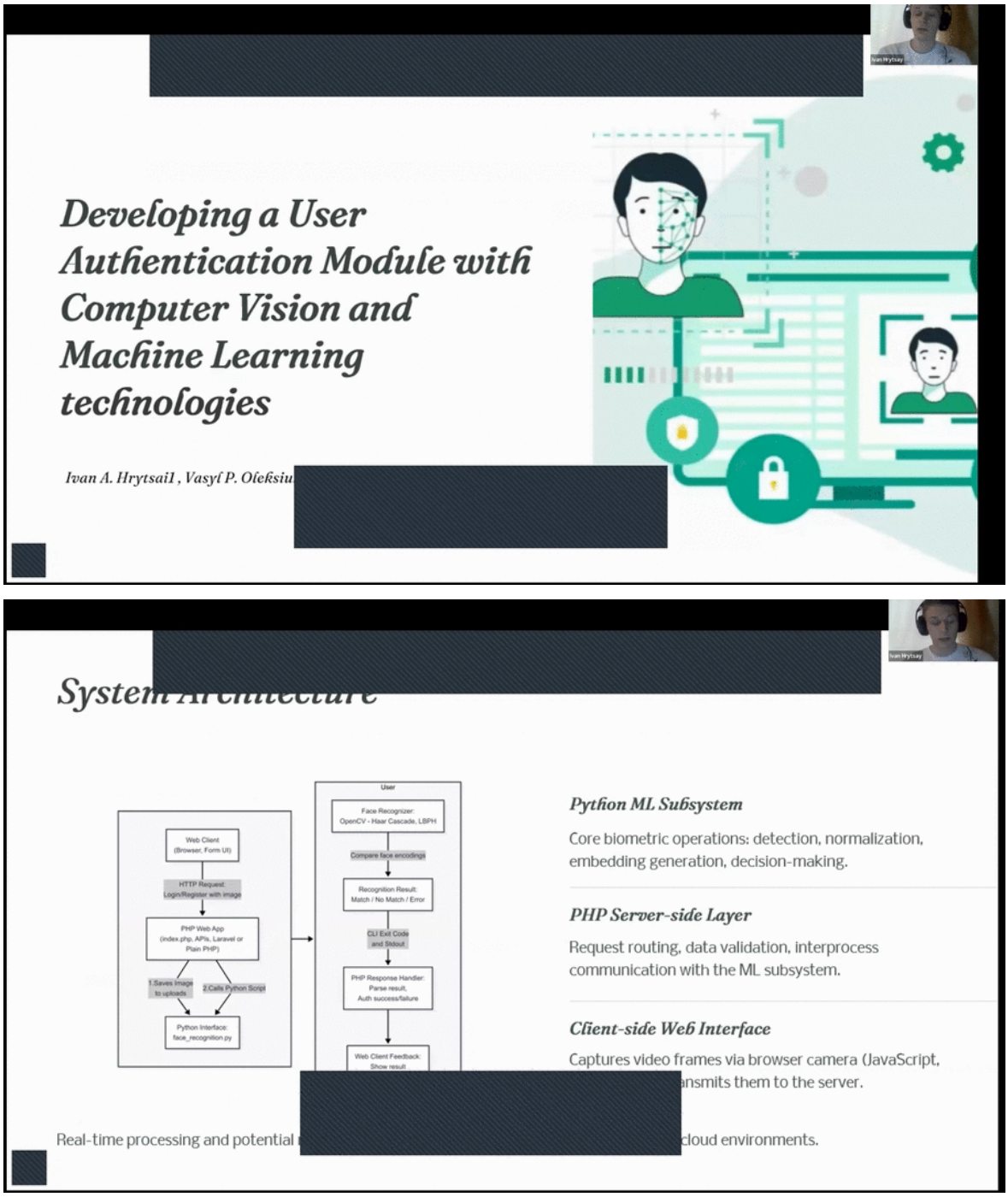


Figure 19: Excerpts from the presentation of Hrytsai et al. [144] (panels A–B; discussion 00 : 20 : 50–00 : 31 : 04).

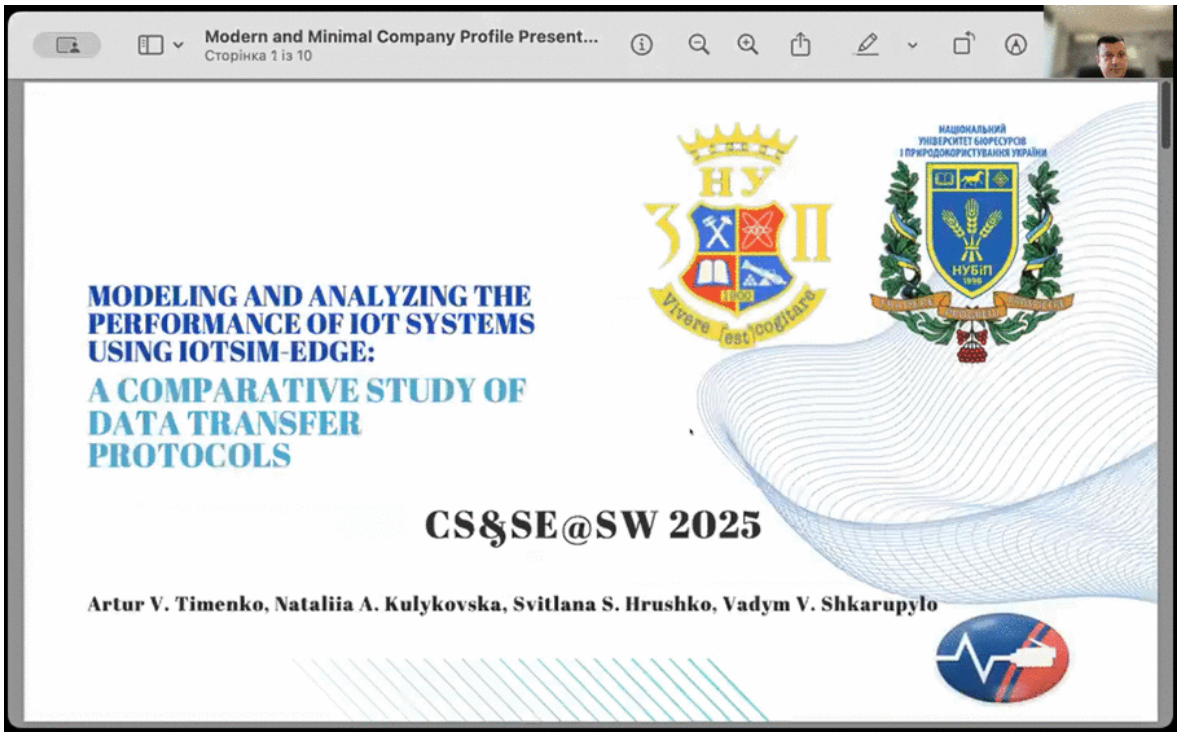
generates 128-dimensional embeddings from a pre-trained ResNet-34, wraps them in a web interface with PHP server logic and a Python imaging subsystem, and, the part that matters most, adds liveness detection through MediaPipe FaceMesh head-movement tracking, because a face recogniser without an anti-spoofing step is not an authentication system. Testing on LFW and CelebA gives above 96.8 % identification accuracy.

Recognising that spoofing is the actual threat model, instead of treating recognition accuracy as the whole problem, is the paper's best instinct.

The evaluation does not follow through on it. Accuracy on LFW and CelebA measures recognition under benign conditions; it says nothing about the liveness check, which is untested against any

presentation attack, a printed photograph, a replayed video, a mask. For an authentication system the operative numbers are the false-accept and false-reject rates at a chosen threshold and the attack-presentation classification error rate, and none is reported. “96.8 % accuracy” is also the wrong summary statistic for access control, where the two error types carry very different costs. The paper is a competent engineering integration whose security claims outrun its testing. It drew the second-heaviest discussion of the workshop, most of it methodological.

Timenko et al. [145] uses IoTSim-Edge to compare MQTT, CoAP and HTTP for edge IoT deployments across response time, delivery reliability and energy, over a twelve-device network (figure 20). The



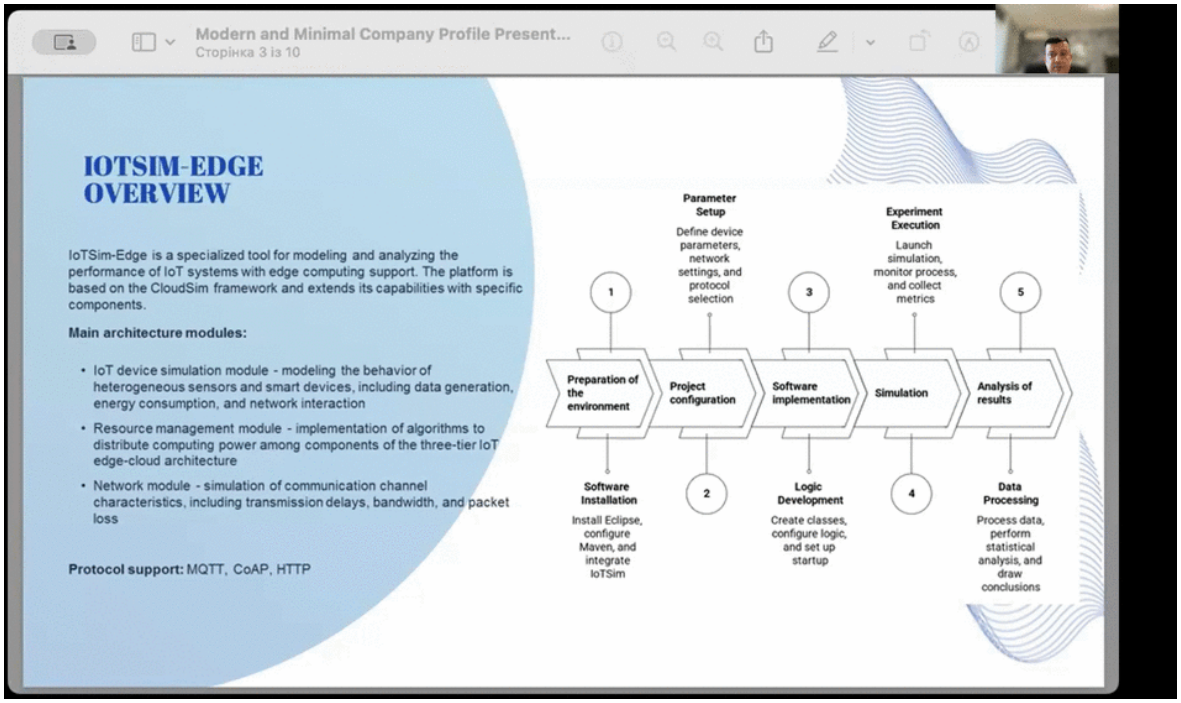


Figure 20: Excerpts from the presentation of Timenko et al. [145] (panels A–B; discussion 00 : 57 : 17–01 : 08 : 46).

results are crisp: MQTT is stable at 42.0 ms with 100 % delivery; CoAP is 40.6 % faster at 22.7 ms with 68.3 % higher throughput and 9.9 % lower energy; HTTP, at roughly 4.91 s, is simply disqualified from real-time use. The paper's secondary claim, that IoTSim-Edge is an effective way to evaluate such infrastructure without buying hardware, is supported by the fact that the study was possible at all.

The value here is decision-useful and cheaply obtained, a designer choosing a protocol for a constrained deployment gets three ranked options with quantified trade-offs, and the two-order-of- magnitude gap for HTTP is the kind of result that settles an argument.

Its weakness is the one every simulation study has and this one does not confront: the numbers are the simulator's, not the network's. No calibration against physical devices is reported, so the sim-to-real discrepancy is unknown, and Kupin et al. [157], in the same volume, identifies precisely that unreported discrepancy as the central gap in simulation-based work. The delivery-reliability comparison is also incomplete: CoAP over UDP and MQTT over TCP differ in their loss behaviour by construction, and reporting 100 % delivery for MQTT without the QoS level makes the comparison hard to interpret. Twelve devices is a small network for conclusions about scalability. This was the most discussed paper of the workshop, eight asks, including four methodological challenges and the workshop's only two questions from the floor beyond one other round.

Zolotopupov et al. [146] implements Kyber (the lattice-based, MLWE-hard key-encapsulation mechanism now standardised for post-quantum key exchange) as a custom .NET library, and takes it further than most implementation papers: the full keygen, encapsulate and decapsulate cycle is reproduced, verified functionally, then benchmarked across parameter sets in both single- and multi-threaded configurations, and finally demonstrated cross-platform on mobile through a .NET MAUI application (figure 21).

Carrying an implementation from theory through benchmarking to a mobile target is more than a tutorial exercise, and the choice of platform is well judged: .NET is short of accessible post-quantum primitives, and a managed-code implementation with published parameter-set benchmarks is genuinely useful to practitioners.

Two cautions belong on the record, and the first is important. A hand-rolled implementation of a cryptographic primitive is a security liability unless it is constant-time and side-channel-audited, and neither property is claimed or tested here; the paper reports functional correctness and throughput, which is necessary and nowhere near sufficient for deployment. Nor are the benchmarks placed against the reference implementation or an established library, so the reader cannot tell whether the performance is competitive. The paper is best read as a study of implementing Kyber on .NET rather than as a library to adopt, and saying so is not a criticism of the work but of how such work is often used.

Zemlianukhina et al. [142] inverts the usual practice in chaotic-system design (figure 22). Instead of exploring a known system's parameter space until an interesting attractor appears, it starts from a desired attractor, expressed as an algebraic curve, and derives a dynamical system that produces it: state variables are tied to the attractor's coordinates by algebraic dependencies, and differentiating those with respect to time yields the equations of motion, with separate and interconnected channels falling out of the coupling. The demonstration builds a system from the Mackey–Glass equation with a cardioid-like polar attractor and realises it on an Arduino Due, where the study also finds hidden chaos, that is, attracting behaviour with no unstable equilibrium to signpost it.

The synthesis-from-specification framing is the contribution and it is a good one, generalising beyond polar curves to any coordinate system in which the target can be written algebraically. Getting it onto commodity hardware instead of leaving it in simulation matters for the claim that these systems can be built with ordinary analog and digital elements.

The evidence is qualitative where it needs to be quantitative. Chaos is demonstrated by phase-portrait resemblance and not by a computed Lyapunov exponent, bifurcation diagram or 0–1 test, and hidden chaos in particular is a claim that wants a basin-of-attraction analysis. With 38 figures and 13 references the paper is illustration-rich and citation-poor, which for a claim of novelty in a well-populated field is the wrong balance; the reader cannot easily see what is new relative to existing attractor-shaping work. Nor is the obvious application, chaotic cryptography or secure communication, evaluated for the

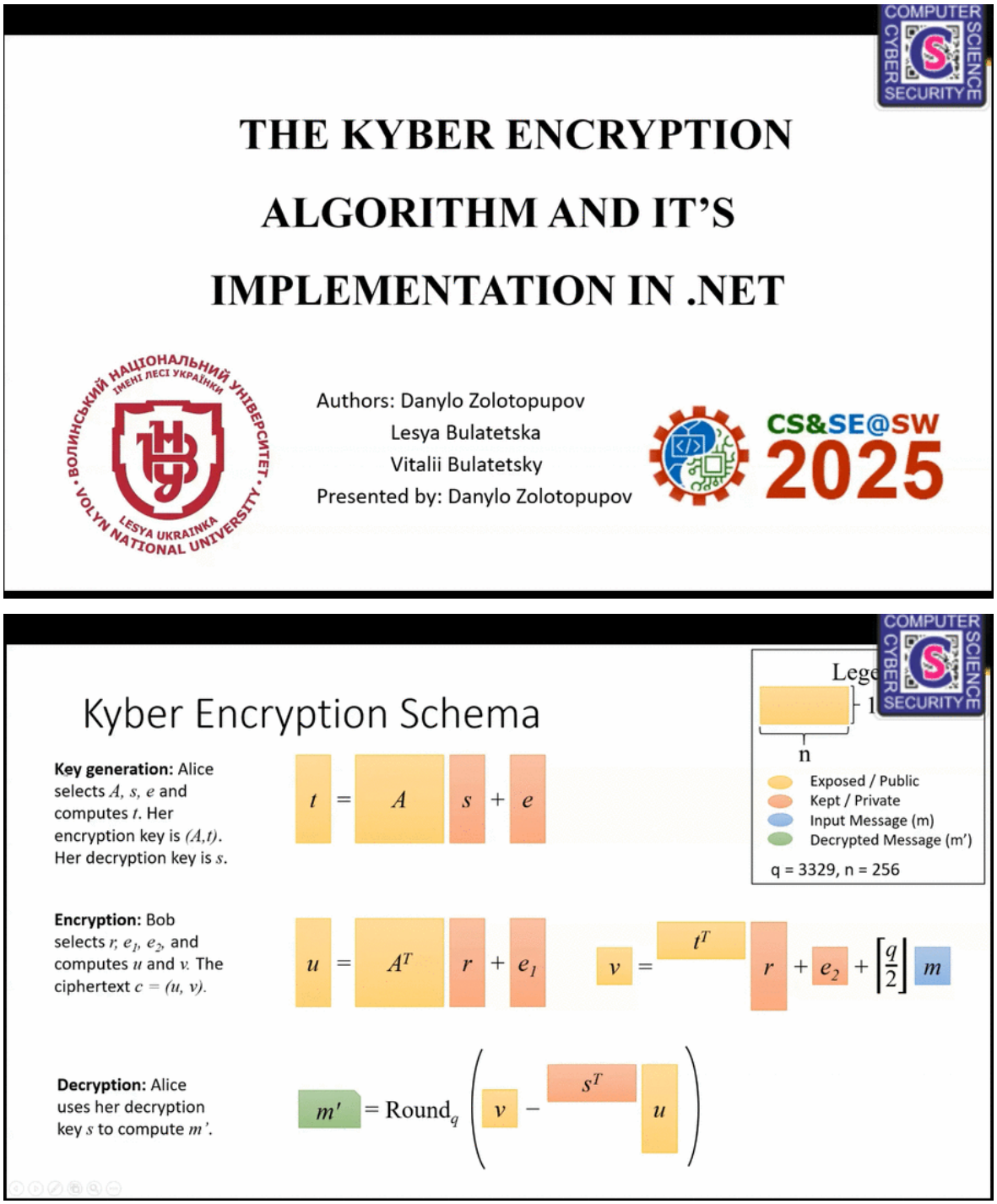
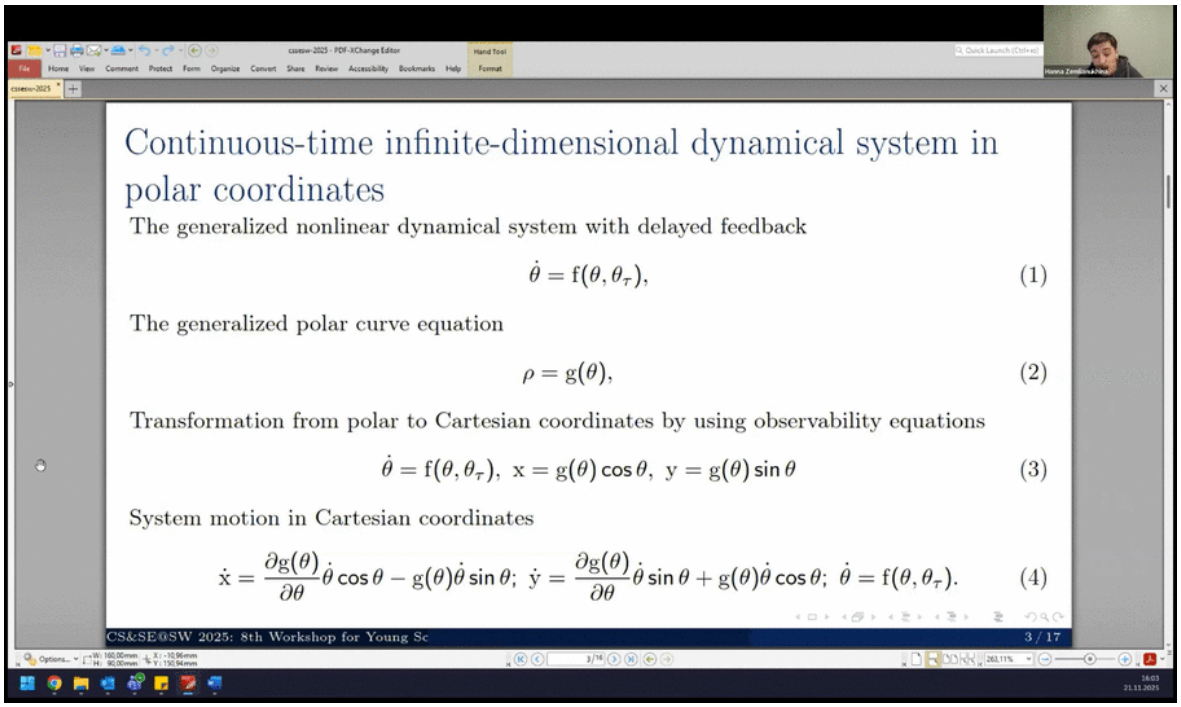


Figure 21: Excerpts from the presentation of Zolotopupov et al. [146] (panels A–B; discussion 05:30:04–05:32:16).

statistical properties it would require.

Semerikov et al. [143] reviews neuromorphic computing for cognitive radio, and its organising claim is a latency argument: sixth-generation wireless needs sub-millisecond spectrum decisions, conventional AI-based cognitive radio sits at millisecond scale, and that gap is architectural rather than a matter of optimisation (figure 23). Spiking neural networks on Loihi, TrueNorth and SpiNNaker are surveyed against spectrum sensing, dynamic allocation and interference mitigation, with reported advantages of sub-millisecond decision latency, two orders of magnitude in energy efficiency, and better noise resilience in RF environments. The paper identifies its own field's gaps (hardware and software



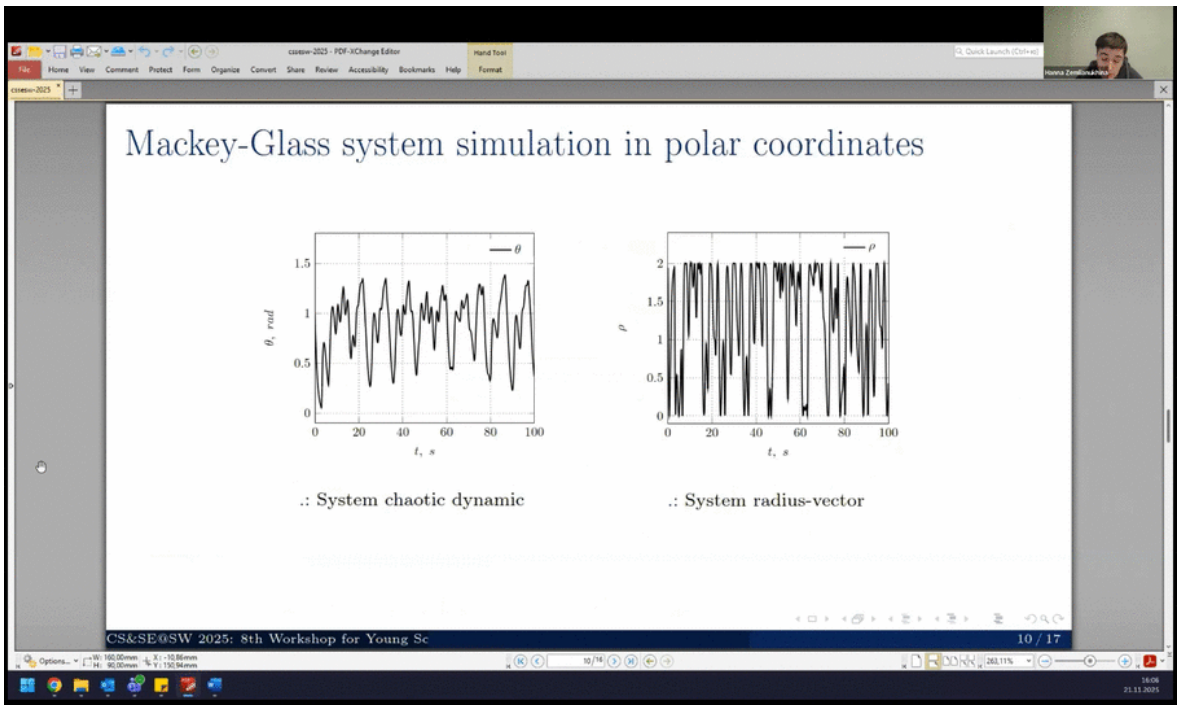


Figure 22: Excerpts from the presentation of Zemlianukhina et al. [142] (panels A–B; discussion 02 : 13 : 46–02 : 19 : 18).

co-design, standardisation, deployment strategy), not only cataloguing successes.

Framing the problem as a latency–energy product and not as accuracy is the right frame for this application, and it is what connects the paper to Timenko et al. [145] and Krak et al. [159]: three papers in this volume hit a hard real-time budget from three unrelated directions (§5.5).

As a review it is narrative rather than systematic: no search protocol, no inclusion criteria, no count of what was screened, so its coverage cannot be assessed and the "two orders of magnitude" figure is inherited from the surveyed sources, not derived under common conditions. Claims of that size need the measurement conditions attached, since neuromorphic energy comparisons are notoriously

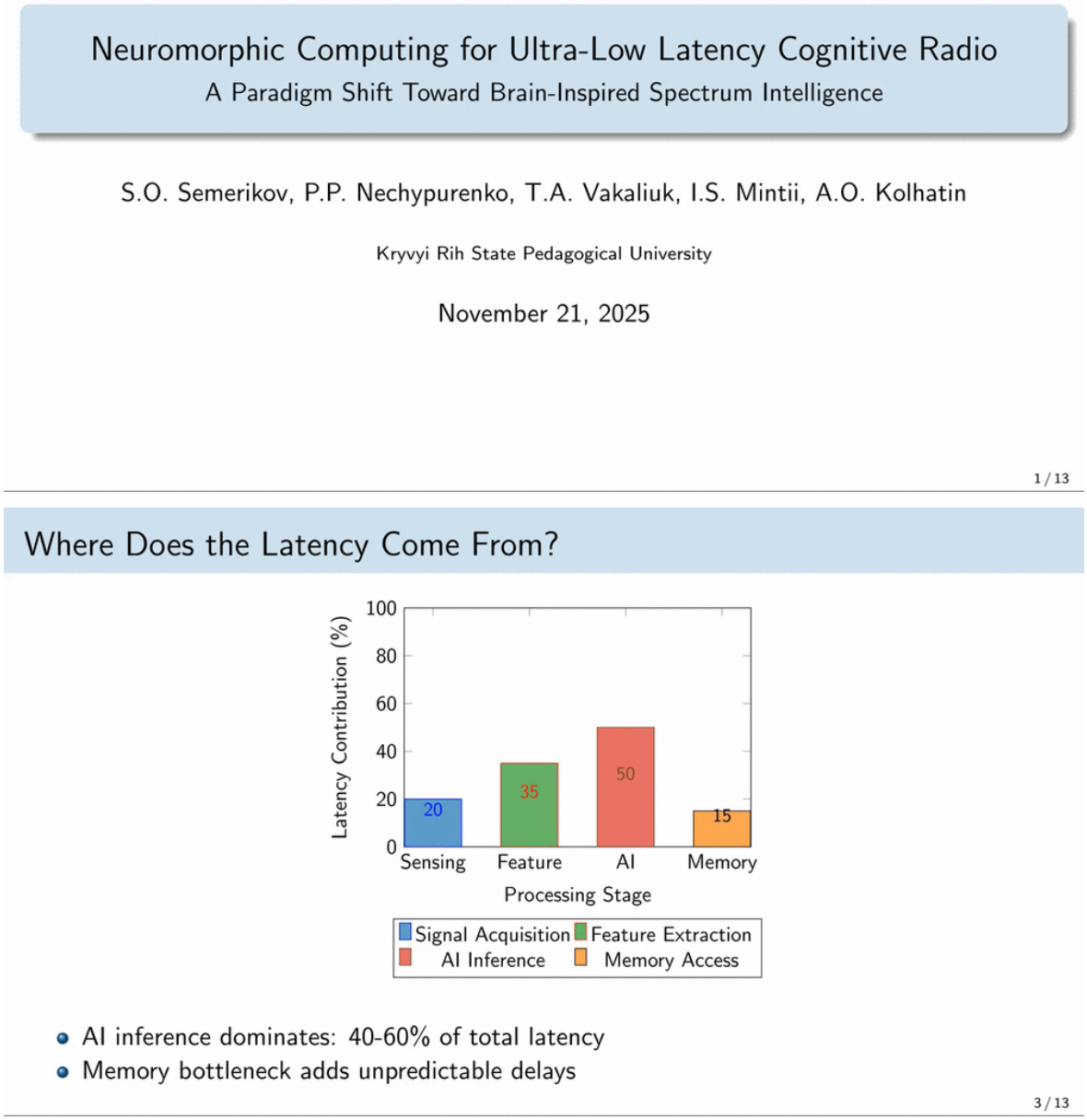


Figure 23: Excerpts from the presentation of Semerikov et al. [143] (panels A–B; presented from a pre-recording, so no live discussion followed).

sensitive to what is counted on the conventional side. There is also no independent experimental contribution. That is reasonable for a review, but it means the latency claims rest entirely on others' hardware measurements. The paper was pre-recorded and played back in the penultimate slot, so it has no discussion and its figure is drawn from the recorded slides.

6.5. Interactive graphics, virtual reality and perceptual interfaces

Tiahunova et al. [147] asks what a game engine costs in hardware terms, and answers it by building a 3D platformer in Unreal Engine 5.4.4 with Blueprint visual scripting and measuring frame rate, GPU utilisation, CPU utilisation and the effect of graphical settings, having first surveyed Unity, Unreal, Godot and Redot on qualitative characteristics (figure 24). The paper is explicit about the project it constructs (genre, target audience, platform, difficulty, expected play duration) which is unusual and welcome, because engine performance is meaningless without saying what was run.

The question is a real one for small teams choosing an engine on a hardware budget, and measuring

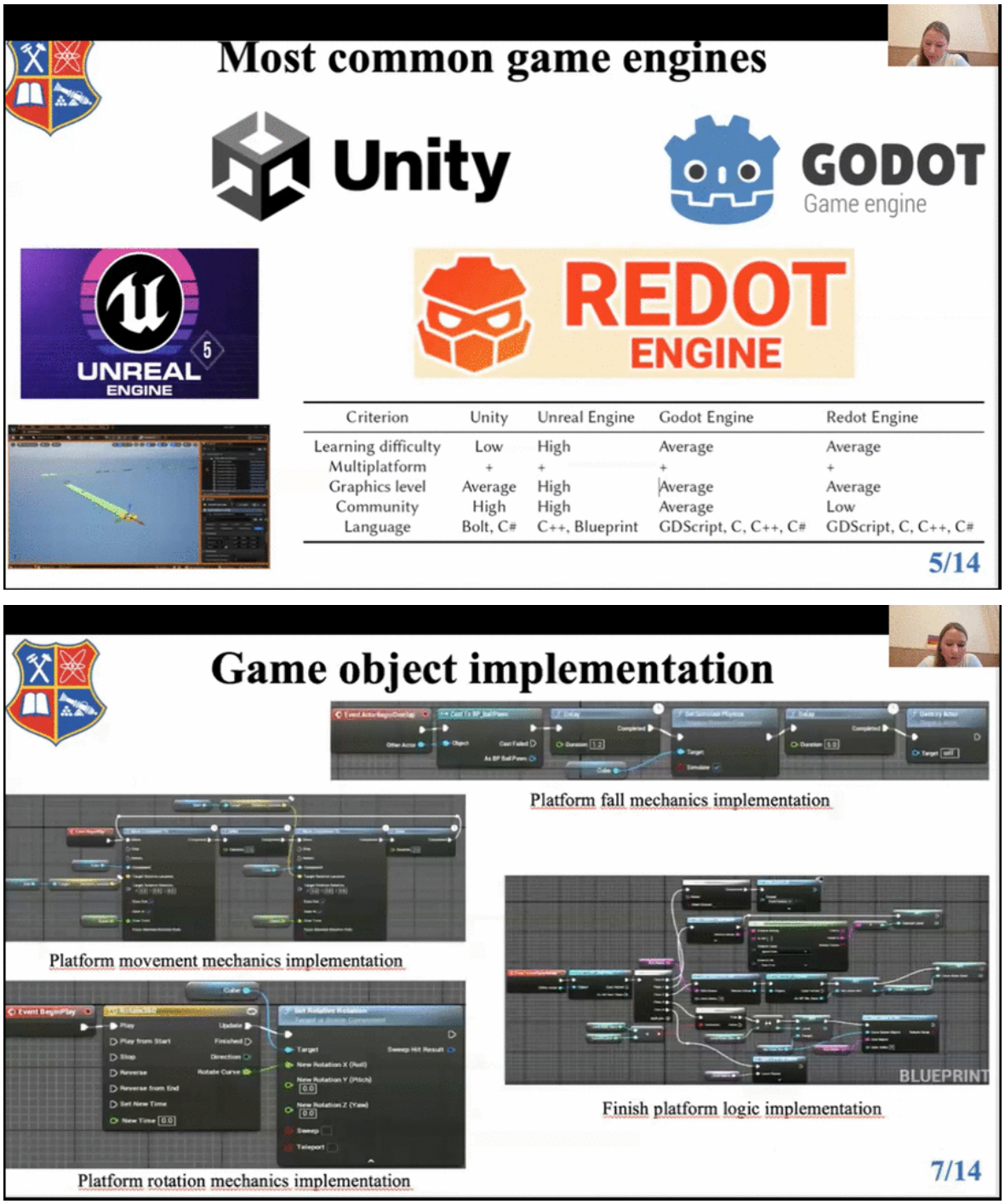


Figure 24: Excerpts from the presentation of Tiahunova et al. [147] (panels A–B; discussion 01:23:57–01:29:22).

it on one's own project and not on a synthetic benchmark is the appropriate way to answer it for a given kind of game.

That is also the study's limit, and it is a sharp one. A single project on a single hardware configuration cannot separate the engine's cost from the project's implementation, and the comparison across engines is qualitative while only the Unreal build is measured, so the paper supports statements about *this game in Unreal* much more strongly than the engine-comparison framing implies. There is no repeated-run variance, no scene-complexity control, and no report of the test machine's specification alongside the numbers. Symmetrical implementations in at least two engines, or a stated identical scene, would have

converted an interesting measurement into a comparison.

This paper is textually the closest neighbour of Brilov and Striuk [148] in the entire volume, at cosine 0.662 (§5.1), despite sitting in a different a-priori cluster and sharing no keywords with it. Its discussion was among the workshop's most methodological, and drew one of the few questions from the floor.

Doroshuk [161] is the shortest and quietest paper in the volume and among the most reusable (figure 25). It takes the practical question of whether two interface colours are *perceptibly* different, and answers it with colour-difference formulae and an explicit colour-discrimination threshold rather than with a contrast ratio, extending the treatment to users with colour-vision deficiency and delivering

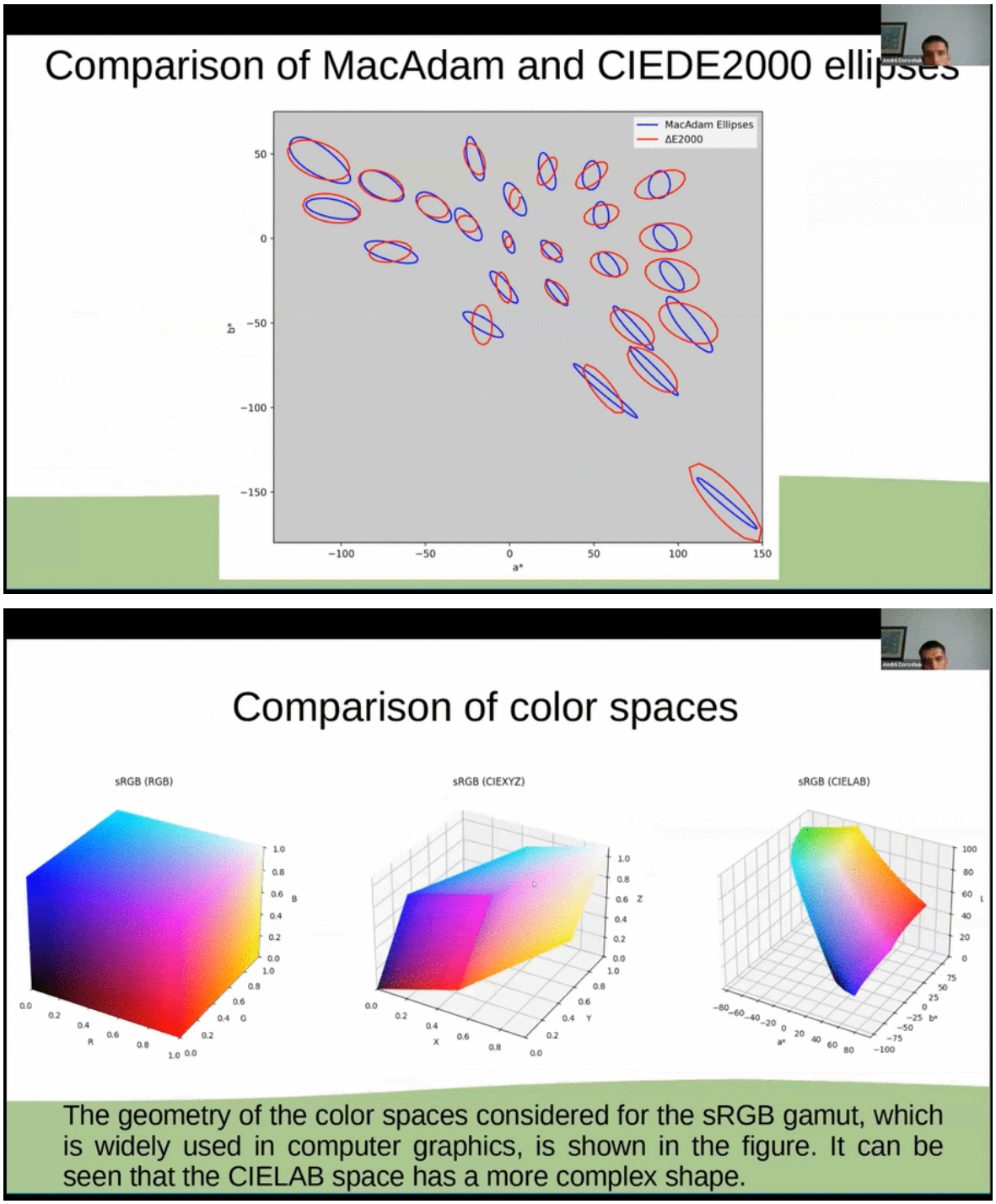


Figure 25: Excerpts from the presentation of Doroshuk [161] (panels A–B; discussion 02:43:53–02:45:42).

software that computes the thresholds. The comparison against the Le Courier table and against contrast ratio is the right validation move for a proposed metric.

The contribution is a decision rule a designer can apply: not "these colours contrast" but "this difference is above or below the threshold at which a human, including a human with anomalous trichromacy, can tell them apart", which is the question interface design actually faces, and which matters more in VR and AR where luminance conditions vary. A single-author paper that produces a usable rule is a good outcome for a workshop of this kind.

What is missing is human data. The paper reasons from established colour-difference formulae to interface recommendations without a perceptual study of its own, so the thresholds are inherited rather than validated in the interface context, and no user experiment tests whether the metric predicts discrimination better than contrast ratio does for real interface elements. The colour-vision-deficiency treatment is a simulation-based transform rather than a study with affected users. The paper is a well-founded proposal awaiting an experiment.

It drew the shortest discussion of the workshop, 1.8 minutes against an allocated minimum of five. The brevity is a scheduling artefact and not a verdict on the paper.

Sinkevych et al. [162] specifies a reproducible pipeline for animated agents in VR: skeletal rigging and keyframe animation in Blender, runtime control through Unity's XR Interaction Toolkit, a scene formalised as agents, interactive objects and spatial trigger zones, and character behaviour driven by a finite-state machine over baked animation clips, with user input, collisions and physics routed through XR controllers (figure 26). A low-polygon VR shooter demonstrates the whole path end to end.

The paper's audience is stated and its design follows from it: educational and small-team research settings, where the obstacle is not novel technique but the absence of a documented path from asset to interactive agent. Formalising the scene model and committing to a component-based, modular architecture is what makes the pipeline extensible toward training simulators, and the finite-state-machine-over-baked-clips choice is a sensible, cheap one for that audience.

As research the contribution is methodological instead of empirical, and the paper does not claim otherwise, but the claim it does make, that the pipeline is reproducible, is the one thing not evidenced: no build times, no asset counts, no report of another team following it, and no released project or repository. For a paper whose value is reproducibility, publishing the prototype would have been worth more than any measurement. There is also no evaluation of the agents as agents: no user study, no comparison against alternative animation approaches, no performance figures for the VR target, which matters because dropped frames in VR are a comfort and safety problem and not an aesthetic one.

6.6. Computing education and disciplinary reflection

Bodnenko et al. [149] makes the case for VBA macros as a teaching instrument in science and mathematics – for didactic visualisation, for automating computation, and for building interactive interfaces in which a student can run a controlled virtual experiment (figure 27). The methodological progression it proposes, from a basic model of a phenomenon to an interactive interface over it, is a sound piece of instructional design, and the paper is unusually attentive to the practical reality that much school and university material already exists as MS Office documents produced by older software.

Its strongest evidence is a teacher-training programme, and the finding that matters is that the approach worked for teachers with no prior programming experience. That is the claim on which the whole proposal stands or falls, and it is why this is one of only 2 papers in the volume to reach the second-highest maturity level: something was delivered to real practitioners and observed.

The reporting of that programme is where the paper falls short of its own result. Participant numbers, duration, instrument and outcome measure are not given in the detail needed to evaluate "confirms the effectiveness", and there is no comparison group, so the finding is a well-motivated professional-development report, not a measured effect. The paper's own proposed future work (assessing the impact on students' critical and algorithmic thinking, and on mathematical modelling ability) is precisely the study that would establish the case.

A separate observation belongs to the technology choice instead of the pedagogy: VBA is a declining

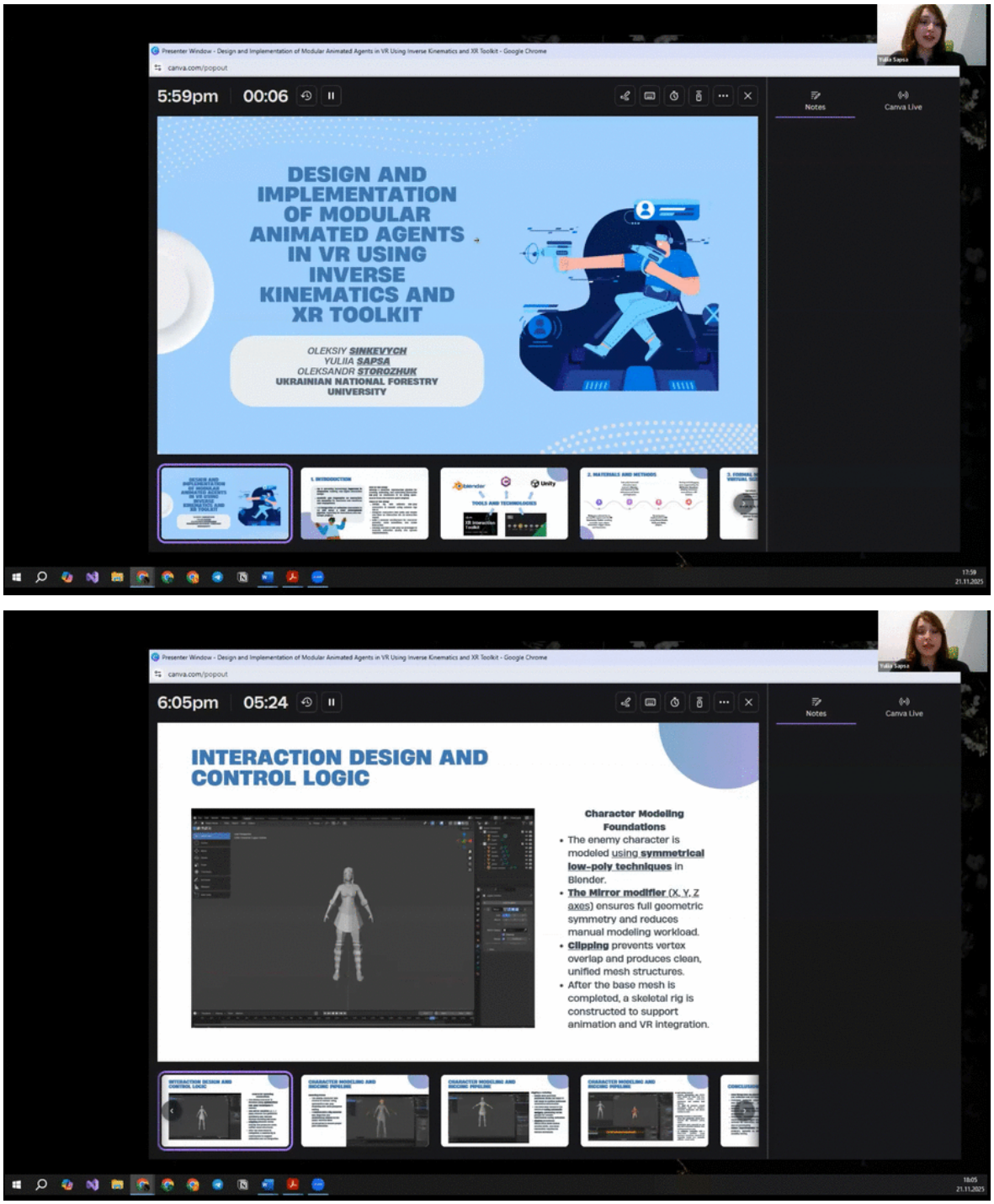


Figure 26: Excerpts from the presentation of Sinkevych et al. [162] (panels A–B; discussion 04:15:18–04:19:06).

platform with a constrained security posture in managed environments, and the paper does not weigh that against the considerable advantage that it is already installed everywhere its audience works. The trade-off is real and arguable in both directions; it deserves a paragraph. One ask in discussion, and this is the round whose captions render in Cyrillic orthography (§2.5).

Brilov and Striuk [148] builds a visual novel about the 1968 NATO Software Engineering conference and the software crisis, in Ren'Py, cross-platform to web, Windows, Linux, iOS and Android (figure 28). The construction is the interesting part: characters are identified from the conference report itself, their dialogue is built from the theses of their recorded contributions, and locations and portraits are

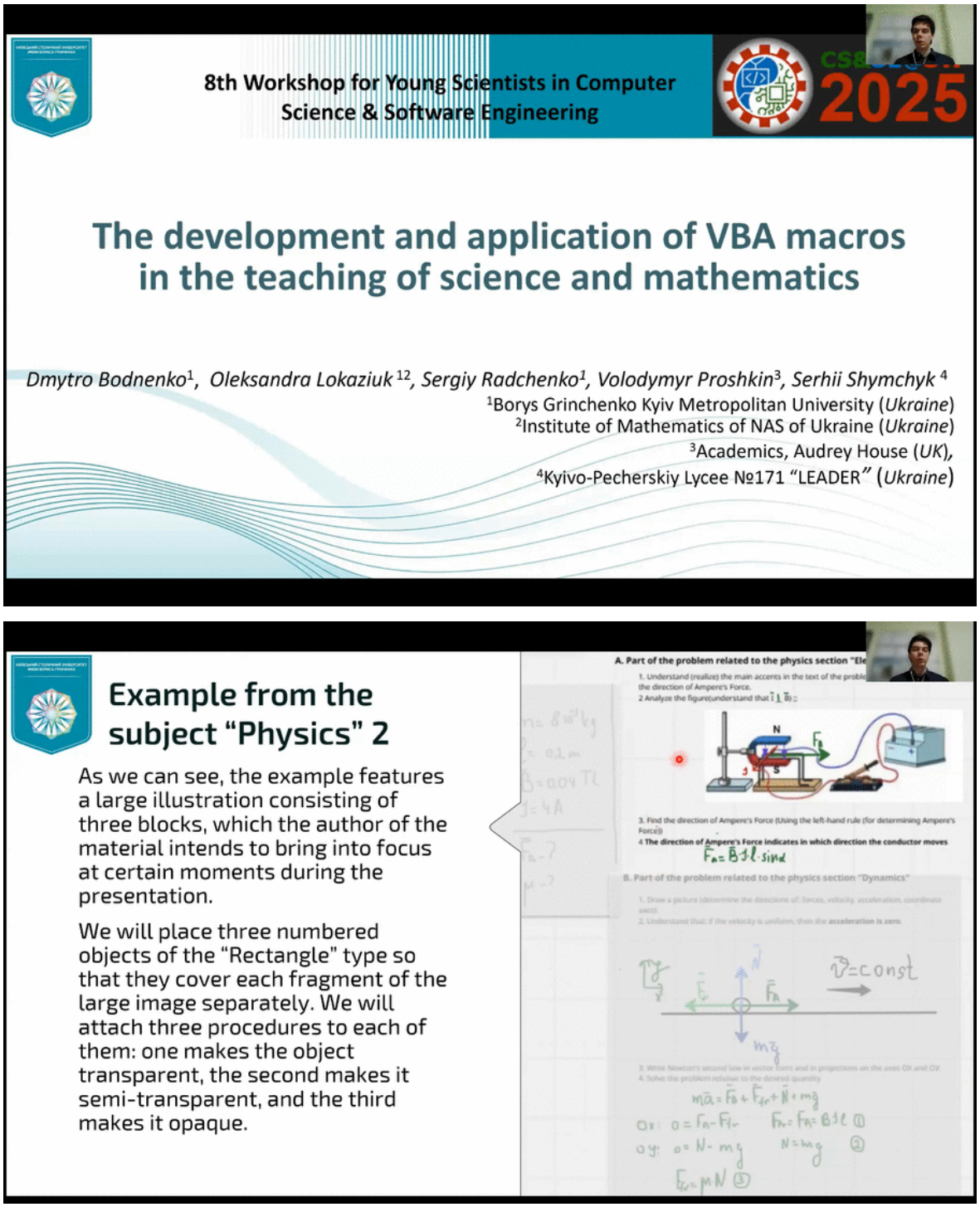


Figure 27: Excerpts from the presentation of Bodnenko et al. [149] (panels A–B; discussion 02:31:08–02:33:48).

generated from period photographs with AI. The genre choice is argued from an analysis of comparable products rather than assumed, and the stated purpose, an illustrative supplement to a Fundamentals of Software Engineering course, is modest and achievable.

Grounding the script in the primary document is what raises this above a themed quiz: the discipline's founding arguments are put in the mouths of the people who made them, taken from what they actually said. As disciplinary self-reflection in a volume that otherwise builds systems, it is a distinctive contribution, and at 8 pages and 6 references it is also the volume's shortest and most lightly cited

Figure 28: Excerpts from the presentation of Brilov and Striuk [148] (panels A–B; discussion 05:47:05–05:52:07).

paper, the latter is a genuine weakness for a historical argument, which needs historiography as much as it needs a primary source.

There is no evaluation. No student has used it in the reported work, so the educational claim is untested; there is no learning measure, no usability data and no comparison against reading the report. The AI-generated depiction of identifiable historical figures also raises a question the paper does not address, and it is exactly the kind of question this volume's own AI declarations are supposed to surface: what the boundary is between illustrating a historical scene and putting invented words or appearances on named people. The paper's dialogue is sourced from the record, which is the right safeguard, and

Table 5
The nineteen discussion rounds, in delivered order. Talk, discussion and handover are separated: folding the handover into the discussion overstates it by about a tenth. The acceptance letter announced 7–10 min of talk and 5–7 min of discussion.

		plan	talk	Q&A	hand	words	turns	asks	chair	floor	advice
1	[144]	2	10.53	10.23	1.15	771	4	6	6	0	1
2	[5]	3	5.52	4.18	0.32	275	1	1	1	0	0
3	[145]	5	14.55	11.48	1.65	961	8	8	6	2	2
4	[147]	4	14.08	5.42	1.1	507	3	5	4	1	0
5	[152]	6	10.72	8.8	0.63	861	3	4	4	0	0
6	[156]	7	5.65	8.52	0.5	737	3	4	4	0	1
7	[142]	8	9.02	5.53	0.57	451	3	4	4	0	1
8	[149]	9	11.12	2.67	0.72	110	1	1	1	0	0
9	[161]	10	9.32	1.82	0.77	116	2	2	2	0	0
10	[151]	11	17.33	4.28	1.68	319	1	1	1	0	0
11	[150]	12	6.4	6.72	0.38	532	4	4	4	0	1
12	[153]	13	13.37	11.15	0.37	989	5	5	5	0	2
13	[157]	14	11.88	6.02	0.48	482	2	2	2	0	1
14	[162]	15	8.78	3.8	0.75	367	2	2	2	0	0
15	[158]	16	24.1	6.37	0.38	508	2	2	2	0	1
16	[159]	17	9.33	3.87	0.35	312	2	2	2	0	0
17	[160]	18	7.45	3.53	0.72	281	2	2	2	0	0
18	[146]	19	14.07	2.2	0.8	174	1	1	1	0	0
19	[148]	20	13.65	5.03	1.17	339	2	3	0	3	0
total			216.9	111.6	14.5	9092	51	59	53	6	10

saying so explicitly in the paper would cost one sentence.

Its discussion drew two of the workshop's 6 questions from the floor, the highest of any round, one of them grounded in the questioner's own experience of the material. Textually it is the twin of Tiahunova et al. [147] (§5.1), which no keyword in either paper would ever have revealed.

7. What the discussion revealed

7.1. Anatomy of nineteen discussion rounds

19 of the 23 papers had a live discussion. One was pre-recorded and one presented without its authors in the room; both fall outside every denominator in this section, since including papers that were not discussed would measure something else. The 19 rounds comprise 758 caption cues and 9,092 words.

Table 5 gives each round in delivered order with its three spans separated into handover, presentation and question time. The separation matters quantitatively. Across the session, handover accounts for 14.5 minutes or 4.2 % of in-round time, presentation for 216.9 minutes or 63.2 %, and question time for 111.6 minutes or 32.5 %. Folding the chair's announcement and screen-sharing into the discussion, as a two-span reading would, reports discussion at 36.7 % instead of 32.5 %. A tenth of the quantity being measured thus depends on where a single boundary is drawn, which is why the protocol in §2.5 requires three spans.

The published schedule and the delivered one are different documents. Figure 29 plots both against the acceptance letter's allocation of 7–10 minutes of talk and 5–7 of discussion. Talks ran from 5.5 to 24.1 minutes with a mean of 11.4, and only 5 of 19 fell inside the allocated window; question time ran from 1.8 to 11.5 minutes with a mean of 5.9, with 6 of 19 inside its band. The longest talk was almost two and a half times its allocation, and the shortest question period was a third of its floor.

The direction of the effect is consistent: substantive talks were allowed to run and the time came out of discussion. The same trade has a seven-year history in this series, the first edition having allocated

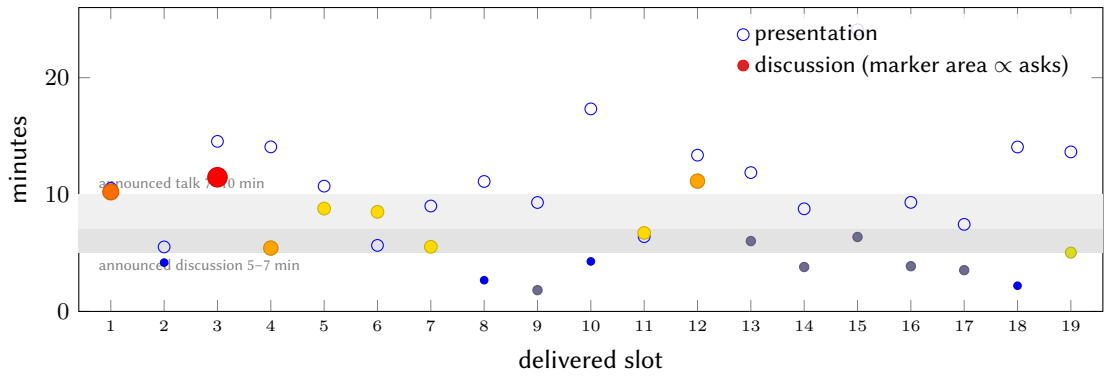


Figure 29: Talk and question time per round against the allocation announced in the acceptance letter (7–10 min of talk, 5–7 of discussion), with the number of asks each round drew. Only 5 of 19 talks and 6 of 19 discussions fell inside their bands. The chairs let substantive talks run and question time absorbed the difference.

fifteen minutes of presentation and seventeen of discussion against this edition's seven to ten and five to seven (§3.1), so that discussion has moved from longer than the talk to roughly half its length.

Boundary confidence is reported per round, not assumed. 15 of the 19 rounds are high-confidence, resting on a presenter hand-off or self-introduction together with a chair handover naming the next title or presenter. The remaining 4 are medium-confidence: two whose question time opens on a chair solicitation instead of a hand-off, one ending inside a stretch of corrupted captions, and one running into the closing address. Every boundary cue is stored with a text fingerprint, and the segmentation script refuses to write its output if a fingerprint ceases to match, so that a re-parse cannot silently shift the table.

One captioning artefact is worth recording, since it resembles a language switch and is not. Eleven of one round's fifteen question-time cues render in Cyrillic orthography while remaining English: the captioner transcribed English speech into Cyrillic letters, so that the chair's "invite next presenter" and the naming of the next speaker both survive as recoverable English spelled in the wrong alphabet. Both phrases transliterate back cleanly, which is how the case was diagnosed, and the same stretch supplied that round's handover evidence. The strings are reproduced in the artefact bundle rather than here, since this document is typeset with a Latin-only font encoding.

7.2. What chairs and floor actually asked

Reading all 19 rounds exhaustively yields 59 asks in 51 turns. The protocol's design estimate was close on asks and 28 % low on turns, since multi-part questions split across more turns than expected.

Reading everything was not diligence for its own sake. The automatic pre-filter's measured recall is 0.56: of 59 asks, 33 overlap a cue the filter proposed and 26 do not, and three entire rounds produced no candidate at all while containing four asks between them. Any study of this kind built on interrogative-marker detection alone would have missed two asks in five and three rounds completely. The figure is recomputed from the candidate list on every run, not quoted from the design.

Two distributions carry the section. The first is who asked. 52 of 59 asks (88 %) came from a chair and 6 from the floor, in 3 rounds only; the remaining 1 sits on a turn whose speaker the transcript does not identify. A workshop for young scientists in which the young scientists asked 6 questions is this section's central and least comfortable result. It is not a failure of the chairs, who asked searching questions; it is a fact about how an online session of this size distributes the right to speak, and it is measured instead of impressionistic. Load per round is correspondingly uneven: one round drew eight asks, another six, two more five each, and three rounds drew one each, with a mean of 3.1.

The second is what was asked about. Figure 30 gives the topic shares, and they are dominated by method rather than by purpose: construct and method validity (19 assignments), missing baseline or comparison (12) and external validity (9) together make 40 of 91 topic assignments. The chairs pressed hardest on whether the evidence supports the claim, which, set against a committee that scored novelty

at 2.43 and everything else generously (§3.2), is a consistent picture of what this community reviews for. Deployment, scale and cost follow at 14 and architecture at 12. By act, information requests (26) and challenges (19) dominate, with recommendations (7) and explicit camera-ready directives (3) behind them, the last of these a category worth having, since the camera-ready deadline fell after the workshop and a chair could still ask for a change.

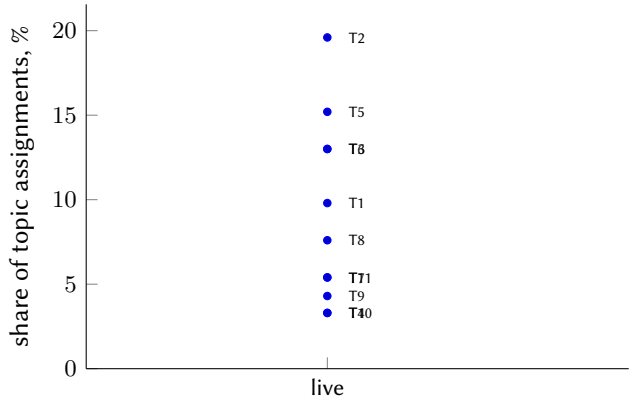


Figure 30: Share of the 91 topic assignments over the 59 live asks. Construct and method validity (T2), missing baseline (T3) and external validity (T1) together account for 40 of 91: the chairs pressed hardest on whether the evidence supports the claim. T9 and T11 carry fewer than eight positives each and are never reported alone.

Figure 31 places every ask in time within its round, which shows the shape a reader would otherwise have to imagine: asks cluster in the first few minutes of question time and thin out, and the rounds with a single ask are visibly the short ones.

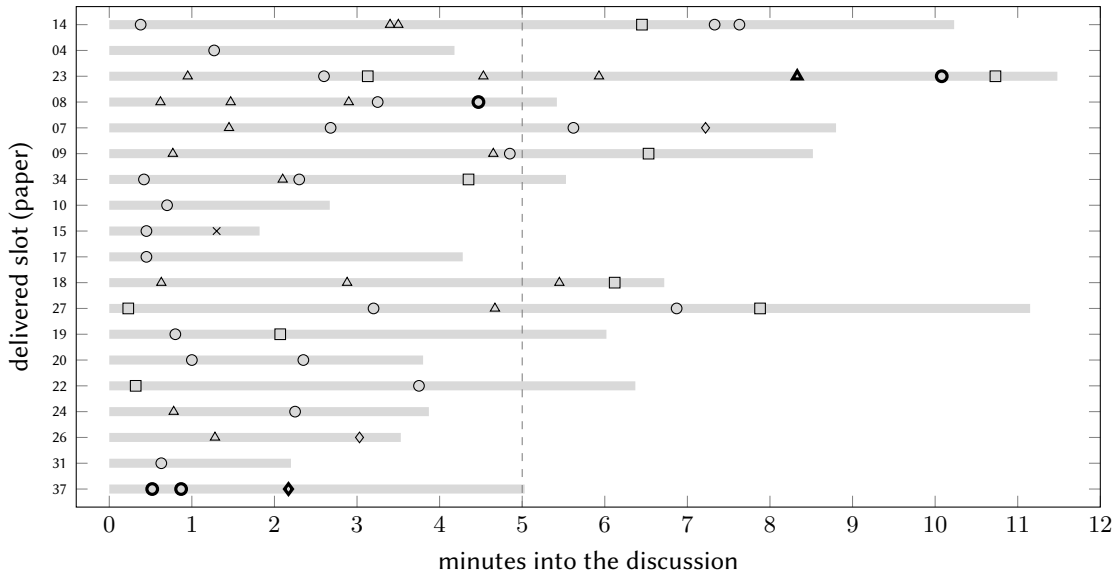


Figure 31: Every ask placed in time within its round, in delivered order. The grey bar is the length of each round's question time. Asks cluster in the opening minutes and thin out; the rounds with a single ask are visibly the short ones.

One methodological event belongs here, not in §2.5, because it is a result about the codebook. In one round a chair says *"Well, another question."* `02:45:11` and no question follows. The codebook required every ask to carry a topic, so the manual pass assigned one. Two of the three model raters refused to assign any topic across four seeds. They were right and the human coder was wrong: an announced-but-never-asked move has no topic because no question was uttered. The topic dimension is now declared partial over the act space, empty for that act and only for it, the manual code was corrected to empty, and the extraction script asserts the rule so it cannot drift back. All three models

then agreed unanimously on the act, the absent topic and the absent response. This is the model panel correcting the human coder, and it is reported as such because it is the strongest evidence in the paper that the audit described in §2.3 does something.

7.3. Prepared and delivered questions

Question sets were generated for each paper with a language model before the session and were available to the chairs while moderating. Because the chairs wrote those questions and had them to hand, the relation between the prepared set and the live discussion is causal rather than predictive, and no claim is made here that a model anticipated the discussion. What can be measured is how much of the prepared material was used. Parsed with the splitting rule applied to the live asks, the prepared material comes to 105 items, of which 95 items and 211 asks belong to the 19 papers that had a discussion, giving 3.58 prepared asks for every live one. Of the live asks, 48 % can be traced to the prepared set, and 82 % of those matched pairs are close paraphrase or near-verbatim recitation, so the material was more often read than reworked. 52 % of live asks have no counterpart in the prepared set. Of the questions asked from the floor, 0 of 6 were anticipated, which is the only figure in this subsection not confounded by the chairs having written what they then read; with $n < 10$ it is exploratory. The prepared sets functioned as a safety net, in that no round went without a question, but they did not substitute for engagement: over half of what was asked was improvised, and everything from the floor was.

7.4. Answers and their limits

This subsection reports a distribution without interpreting it as a measurement. Response coding is the least reliable dimension in the study, at $\alpha = 0.37$ with only 20 % of asks coded unanimously by four raters (§2.5), so the ordering of its categories cannot bear weight.

Descriptively, evidence was supplied in 15 cases, a partial or scope-limited answer given in 12, a commitment to future work in 11 and an assertion without evidence in 10, while 6 answers are not recoverable from the captions and 5 asks received no answer at all. The last two categories are kept distinct because they differ in kind: one is a limitation of the recording, the other an interactional fact about the session. Together they account for 11 of the 59 asks, roughly a fifth, that produced no usable answer.

Set against figure 29, where talks overran and question time absorbed the difference, this is consistent with the trade-off described in §7.1: compressed question time yields fewer complete answers. The single explicit time-pressure closure in the recording falls at the end of a round that had already drawn five asks, which suggests the constraint was the schedule instead of the paper.

8. Positioning against the 2024–2026 literature

8.1. Where the volume is ahead, aligned and lagging

The retrieval protocol of §2.6 asked, for each cluster, what the 2024–2026 literature holds and where this volume's papers sit against it. Verdicts are stated as one of four: *ahead*, where a paper anticipates or independently reproduces a frontier result; *aligned*, where it sits within current practice; *lagging*, where the frontier has moved past its method or its evidence standard; and *orthogonal*, where the comparison does not apply. Table 6 gives them per cluster with the DOI-verified sources.

Before the verdicts, the measurement that frames them. Of the 77 frontier references shortlisted by the protocol, four are cited anywhere in this volume's 453 unique cited DOIs, an overlap of about five per cent on the questions this review asks. That is not a criticism of the authors, whose reference bases are strikingly current (§4.2): it is a statement about how narrow the intersection is between a literature that is current *in a field* and a literature that is current *on a specific question*. Both can be true simultaneously, and the distinction matters for how the result is read.

Table 6
Each cluster against the 2024–2026 literature retrieved by the protocol of §2.6. Verdicts are *ahead*, *aligned*, *lagging* or *orthogonal*, and every one is argued in §8.1 against a source carrying a resolvable DOI. The lags fall almost entirely on evidence standards rather than on technique.

Cluster	Verdict	What the frontier holds, and where the volume sits
C1	**ahead** (1), aligned (3)	Typed, schema-constrained tool interfaces are the direction of travel; gains are in structural validity, not task success [163, 164]. Slobodianiuk and Semerikov [155] independently reproduces the review-level finding that innovation moved off the backbone.
C2	aligned	Multi-sensor fusion and detector–tracker–forecaster pipelines are live frontiers [165, 166]. No shared simulator protocol exists, as Kupin et al. [157] argues, but task-specific suites and sim-to-real reporting do [167].
C3	aligned (method), **lagging** (validation)	Model comparison on tabular and image-derived data is standard practice; single-dataset accuracy is no longer accepted as evidence of robustness.
C4	aligned, **lagging** (one paper)	Protocol ordering confirmed on hardware, with magnitudes shifting once broker behaviour and real energy draw are measured [168, 169]. Face authentication has moved to attack-diversity evaluation; LFW and CelebA no longer evidence operational readiness.
C5	aligned	Perceptual colour metrics supplement, not replace WCAG contrast, and no universal threshold is yet justified [170], which is why Doroshuk [161]'s computed threshold is the right form of answer.
C6	orthogonal	Not competing on method novelty. Judged instead against the evidence standard for educational claims, which both papers state as future work.

Language models as software substrate: ahead on one point, aligned otherwise. The frontier confirms the architectural shift these papers bet on. Since 2024 the tool-using literature has moved decisively toward typed, schema-constrained interfaces, and the measured benefit is concentrated in structural validity and not in end-to-end correctness; schema enforcement and constrained decoding reliably reduce parse and compile failures while leaving semantic grounding an open problem [163, 164, 171]. Zhakhalov [152] and Baran et al. [153] are squarely inside that movement, and §9 treats their convergence in detail. The clearest *ahead* verdict in the volume belongs to Slobodianiuk and Semerikov [155], whose narrative extension finds that the locus of innovation has moved from backbone architecture to post-training, inference-time computation, retrieval and tool use, which is precisely what the review-level literature of 2025–2026 independently concludes [164]. A student review arriving at the same conclusion as the surveys, from a different corpus, is a genuine result and deserves to be named as one. Lytvynenko et al. [154] is aligned but conservative: preprocessing sensitivity is a real and under-reported effect, and lexicon-based scoring is a weaker instrument than the field now uses.

Airborne sensing: aligned, with one lagging evidence standard. The operational papers sit comfortably in current practice. Multi-sensor fusion of SAR with optical imagery for disturbance mapping, and detector–tracker–forecaster pipelines for small aerial objects, are both live frontiers rather than settled ground [165, 166], and Chaskovskyy et al. [158] and Krak et al. [159] report the metrics that literature reports. The specific comparison Chaskovskyy et al. [158] runs, a random forest against a U-Net for forest-cover change from Sentinel imagery, is itself an active question in 2024–2026, not a settled one, with recent studies reaching the same architecture pairing independently [172, 173, 174]. Kupin et al. [157]'s central claim needs one qualification the volume could not have known without this retrieval: the community has indeed *not* converged on a single benchmarking protocol, which is

the paper's finding, but it has built reusable benchmark suites and competition tracks for particular tasks, and sim-to-real discrepancy is now routinely reported task-specifically as trajectory or state error instead of not at all [167]. The paper's diagnosis holds; its implicit premise that nothing exists to build on does not, and a revision should cite those suites.

Applied machine learning: aligned on method, lagging on validation. Comparing tree ensembles against deep networks on tabular and image-derived risk data is exactly what this literature does, and the shallow-versus-deep question for air-quality prediction specifically is open [175, 176, 177]; so is the data-efficiency of vision transformers on small datasets [178, 179, 180]. Rudnichenko et al. [151] and Mazurets et al. [160] run both comparisons competently. The gap is in evidence standards rather than technique: single-dataset accuracy above 98 % is no longer treated as evidence of robustness in the venues these papers aim at, where cross-domain and deployment-oriented evaluation is now expected. That is the same criticism §6.3 makes from inside the papers, and the frontier makes it from outside.

Signals, security and constrained computing: mixed, and the sharpest lag. Timenko et al. [145]'s protocol ordering is confirmed by the recent literature: MQTT for dependable delivery, CoAP for latency in lightweight scenarios, HTTP only where interoperability outweighs efficiency [168, 169], and that is a real validation of a simulation study. The same literature also states the qualification: on real hardware the *size* of each advantage shifts substantially once broker behaviour, connection persistence and physical energy draw are measured, which is why the missing calibration in §6.4 matters instead of being a formality. The volume's clearest lagging verdict is Hrytsai et al. [144]'s evaluation. Since 2024 face authentication has moved from benchmark accuracy on standard datasets toward generalisable presentation-attack detection under attack diversity and deployment constraints; LFW and CelebA remain useful datasets but are no longer accepted as evidence of operational readiness for a system claiming liveness detection. The paper's engineering instinct, add a liveness check, is right and current; its evidence base is two generations behind what the claim now requires. Zolotopupov et al. [146]'s Kyber work sits differently: implementing and benchmarking ML-KEM on a new platform is a recognised contribution in this literature [181, 182], which is why the missing constant-time and side-channel analysis is the gap to close instead of the choice of target.

Human-centred computing: aligned, and correctly cautious. Doroshuk [161]'s reliance on perceptual colour-difference thresholds rather than contrast ratio matches where this literature has arrived: WCAG contrast satisfies formal accessibility minima while perceptual metrics such as CIEDE2000 and CAM16-UCS are the tools for reasoning about actual distinguishability, particularly for users with colour-vision deficiency [170]. Notably, the frontier does *not* yet support a single universal threshold for all interface elements, so the paper is right to deliver a computed threshold and a tool, not a constant, and a revision should say that explicitly. Game-engine comparison is likewise a live genre with its own recent entries [183, 184, 185], and the standing weakness there, comparing engines qualitatively while measuring only one, is exactly the one §6.5 identifies in Tiahunova et al. [147]. Sinkevych et al. [162] is orthogonal to the research frontier by design, being a pipeline paper for a stated audience, and is better judged against the reproducibility standards of §10 than against method novelty.

Computing education and reflection: orthogonal to the frontier. Neither Bodnenko et al. [149] nor Brilov and Striuk [148] competes on method novelty, and the frontier comparison is not the right instrument for them. What the external literature does supply is a standard for the claims they make: both rest on educational effectiveness assertions that would need a measured study, and §6.6 says so.

8.2. What the pattern across verdicts says

Two things generalise beyond this volume.

First, the lags are concentrated in evaluation, not in technique. No paper here uses an architecture or method the field has abandoned; several use evidence standards the field has moved beyond. That is a

much more tractable deficiency than the reverse would be, and it is consistent with what the review scores said independently: novelty was marked strictly and presentation generously (§3.2), while every paper in the volume evaluates something (§5.8). The corpus is not short of empirical intent; it is short of the comparative and cross-domain apparatus that would make its evidence travel.

Second, the frontier itself is moving in the same direction. Across 2024–2026 the clearest disciplinary shift in machine learning, natural language processing, computer vision and systems evaluation is away from proposing new methods and toward shared benchmarks, protocols and reproducibility rules [186, 187, 188]; much of the rest of engineering still prioritises novel algorithms and treats standardisation as future work. This volume sits on both sides of that line, and the paper that states it most explicitly is the one whose entire contribution is that a literature cannot be read together for want of a protocol [157].

Figure 32 plots the two judgements that bear on this against each other for the 20 papers with reviews in the export: the reviewers' own novelty scores, and evidence strength as coded by the model panel, which never saw a review. Across those 20 papers the two are essentially unrelated: Pearson $r = -0.12$ and Spearman $\rho = -0.06$, with mean novelty scores spanning only 1.50–3.33 on a five-point scale. Novelty and evidentiary rigour are close to independent in this corpus, so a reviewer's novelty verdict cannot be read as a proxy for how well a paper supports its claims, in either direction. With $n = 20$ this is an observation and not a finding, and the narrow novelty range is itself part of why: a criterion on which almost every paper scores between 1.5 and 3.3 cannot discriminate much of anything.

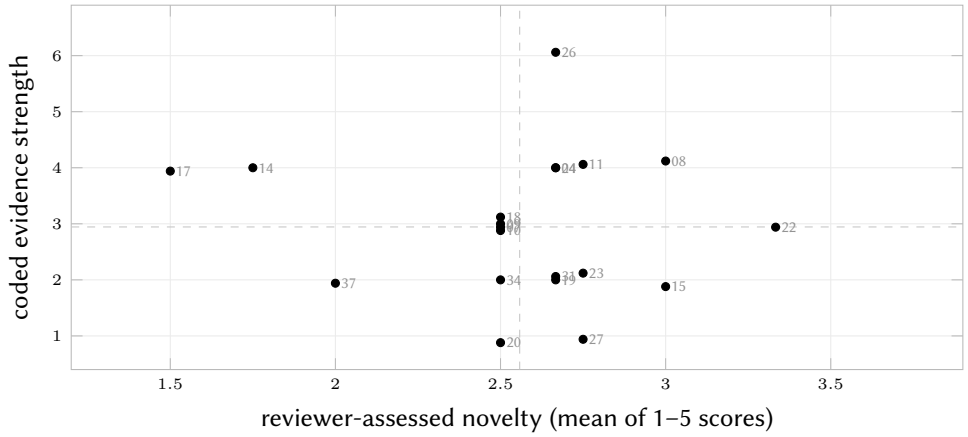


Figure 32: Reviewer-assessed novelty against coded evidence strength for the 20 papers with reviews in the export, labelled by paper number; dashed lines mark the means. The two axes come from independent sources (the reviewers' own scores, and the model panel's coding, which never saw a review) and they are essentially unrelated ($r = -0.12$, $\rho = -0.06$). Three volume papers carry no review in the export and are omitted rather than imputed.

9. The emerging technique: schema-constrained agency

9.1. The convergence, and the evidence for it

Two papers in this volume, from two groups with no shared authors, no shared institution and no shared reference, presented in the same session four talks apart, make the same architectural bet: that the right action space for a language-model agent is a typed schema, and that the schema should be enforced by something other than the model.

Zhakhalov [152] argues it in the general case, for GraphQL: the agent's act step becomes typed query synthesis against a schema, validation happens before execution and returns structured diagnostics, and the schema *is* the sandbox: there is no arbitrary-code path to defend, and injection is confined to a bounded query surface. Baran et al. [153] builds it in a particular case, for Cypher over a product graph,

with Model Context Protocol servers performing the validation so that the model never addresses the database directly.

The evidence that this is a convergence and not a coincidence of topic is quantitative, and comes from §4.4. The cosine similarity between the two papers is 0.650 (rank 2 of all 253 pairs in the volume, against a median of 0.18) and they co-cluster in five of the six representation-and-unit runs. Under two of those runs they form a cluster of exactly two papers and nothing else.

The single run that separates them is the most informative one. It is the sparsest document unit under the sparsest representation: titles and keywords only, term frequencies only. There, the evidence available is two seven-word keyword lists that share not one token. And that is the argument of this section instead of a caveat on it. The two papers share zero references and zero keywords; neither the keyword map nor the coupling graph of §4.3 can see any relation between them at all. The convergence is semantic, and it is invisible to precisely the two instruments a bibliometric review would have used.

A third paper arrives at the same place from a different direction. Slobodianiuk and Semerikov's [155] narrative extension of its own systematic review finds that between 2024 and 2026 the locus of innovation in text generation moved off the backbone architecture and onto post-training, inference-time computation, retrieval and *tools*, the layer where a typed contract lives.

9.2. The claim in falsifiable form

Stated so that it can be tested and, if wrong, shown to be wrong:

> Between 2024 and 2026, engineering effort in agentic systems moved from improving the learned component to constraining the interface between it and a consequential action. The constraint is a machine-checkable contract (a schema, a grammar, a protocol-mediated validator) enforced outside the model. Its measurable benefit is concentrated in structural reliability and in attack-surface reduction, **not** in task success.

The external evidence supports the shift and bounds the benefit tightly. Across 2024–2026, tool-using systems moved toward typed interfaces, structured outputs and validation layers, and the measured gains are in format validity and parse reliability while semantic grounding remains unsolved, since schema compliance can coexist with hallucinated content and with within-scope harmful actions [163, 171, 164]. The last clause is what a reader should hold onto: a validated query can still be the wrong query, and neither paper in this volume claims otherwise, though neither tests it either.

9.3. The hostile reading

A section arguing that a trend is real owes its reader the strongest available case that it is not. There is one, it is DOI-verified, and it cuts deeper than a limitation.

Constraining a model's output can make it reason worse. Tam et al. [189] show that restricting generation to a structured format significantly degrades reasoning ability, and that stricter constraints degrade it further, the effect is on reasoning quality, not merely on formatting. and Schall and de Melo [190] reach the same conclusion under the name of a hidden cost of structure, and add the boundary condition that matters here: the penalty falls hardest on instruction-tuned models, which is what an agent is built from.

So the position of Zhakhalov [152] and Baran et al. [153] is not free. The synthesis is therefore conditional: *typed constraints pay when they encode the task's real invariants and when the intermediate reasoning is left unconstrained; they cost when the grammar becomes a syntactic straitjacket over the model's whole generation path.* A schema over the final tool call, which is what both papers propose, sits on the favourable side of that boundary, because it constrains the action and not the deliberation. Neither paper makes that distinction explicitly, and both would be stronger for making it.

This is also where Zhakhalov [152]'s missing evaluation stops being a presentational weakness and becomes the decisive gap. Its token-economy argument is arithmetic on one task; the reliability claim is architectural; the security claim is untested against an adversary. The frontier literature says the

format-validity benefit is real and the reasoning cost is real, and which dominates is an empirical question about a specific task distribution. That experiment is the paper's obvious next contribution, and it is a workshop paper's-worth of work: twenty tasks, a function-calling baseline, an error taxonomy, and a measured injection attempt.

9.4. Where the pattern does not hold, and a secondary trend

Two boundaries bound the claim. First, this is a convergence of *two* papers in a 23-paper volume, observed at one workshop, and its statistical support is a similarity rank and a co-clustering count, not a trend line over time. The paper's claim is that the convergence is real and detectable, not that it is representative.

Second, the pattern is confined to the cluster where an agent acts on a system. It says nothing about the volume's sensing, signals or education papers, and §8 found the disciplinary shift toward protocols and shared evaluation to be uneven across fields for the same reason.

There is a secondary trend, and it deserves a paragraph instead of a section because the evidence is thinner. Three papers hit a hard latency budget from three unrelated directions: spiking networks on neuromorphic hardware for sub-millisecond spectrum decisions [143], a detection-and-forecasting pipeline reporting throughput on CPU as well as GPU because the deployment is real-time [159], and an edge-protocol comparison in which a two-order-of-magnitude response-time difference disqualifies one option [145]. None cites another; none shares a method. What they share is that *latency is the constraint that decides the design*, and in all three the accuracy question is subordinate to it. That is the same structural move as schema-constrained agency, a non-functional requirement promoted to an architectural determinant, which is why it is worth recording, and it is the pattern most likely to be worth testing on the next volume. Figure 33 contrasts the two architectures and marks where each paper sits.

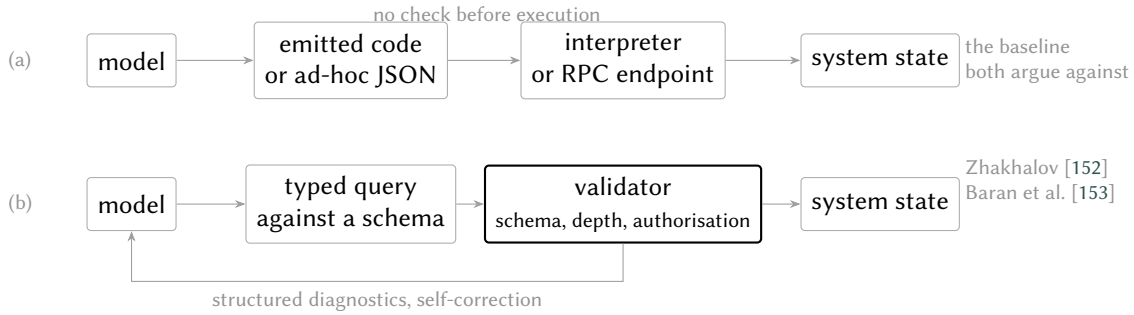


Figure 33: Free-form code-generating agency (a) against schema-constrained agency (b). In (a) the model's output reaches an interpreter or endpoint with no machine-checkable contract in between, so the action space is whatever the runtime permits. In (b) the action is a typed query, a validator outside the model accepts or rejects it before execution, and rejection returns structured diagnostics the model can act on. The measurable benefit is structural reliability and a bounded attack surface; §9.3 gives the DOI-verified evidence that constraining generation can also cost reasoning quality, and why constraining the *action* rather than the *deliberation* sits on the favourable side of that trade.

10. Research agenda and open problems

A research agenda assembled by the editors of a volume they have just read will tend to reproduce the editors' own interests. The construction below is designed to constrain that: each problem must be supported by at least two independent streams of evidence, and the items discarded for want of a second stream are reported alongside those retained.

10.1. Triangulation

Three streams of evidence were collected independently, and a problem is listed only if at least two of them support it.

A – the authors' own future work. 32 future-work sentences extracted mechanically from the conclusions and limitations of all 23 papers, not selected by hand.

D – what the discussion asked for. Asks coded as a recommendation or a camera-ready directive, or carrying the extension topic, from the 59 coded asks of §7.2.

E – what the literature says is open. The 2024–2026 retrieval of §2.6, represented by the DOI-verified entry carrying the point.

23 candidate problems were grouped from those three streams. 17 survive the two-stream filter, 9 carried by all three streams and 8 by two, and 6 are carried by a single stream and excluded. Table 7 lists the survivors with their evidence columns and, for each, what would count as progress; figure 34 shows the provenance, including the candidates that were discarded.

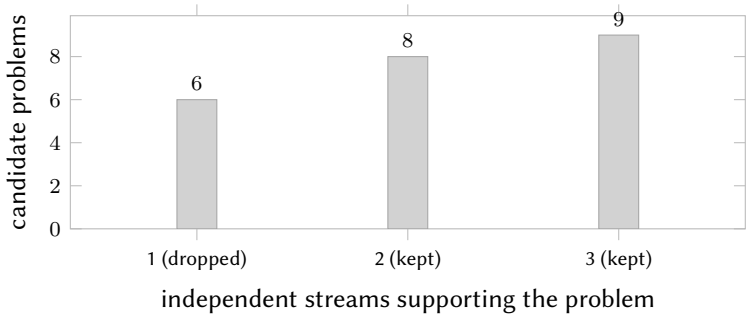


Figure 34: Provenance of the 23 candidate open problems. 9 are supported by all three independent streams and 8 by two; the 6 carried by a single stream are excluded from table 7. The discarded column is shown so that the filter's rejections can be assessed alongside its selections.

Two properties of this construction are editorial and are declared, not buried. The grouping of raw evidence into candidate problems is a human judgement; only the filter is mechanical. And **stream D is not the definition this study set out with**. The design called for using asks whose *answer* was coded as deflected or as a commitment to future work, but the response dimension measured $\alpha = 0.37$ (§2.5), too low to carry a claim, so the stream was moved onto the *ask*, which is coded reliably. Redefining a stream after seeing its reliability is a genuine route to bias, and the only remedy available is to say so where the result is reported. The script that builds the table asserts that every cited ask exists in the coded set and every cited reference exists in the bibliography, so an item cannot claim evidence that is not there, an assertion that fired on an invented ask reference during drafting, which is the only reason this paragraph can be trusted.

10.2. What the volume needs, in the order it can be delivered

The 17 surviving problems fall into three groups. The first is both the largest and the cheapest to act on.

1. *Evidence standards, deliverable in a revision.* Nine of the seventeen are near-horizon problems about how a result is evidenced rather than about what to build. Report the sim-to-real discrepancy instead of the simulated number alone, pressed by this volume's own meta-review, by an ask, and by the external literature. Validate on data from another source instead of reporting single-dataset accuracy above 98 %. Evaluate liveness against actual presentation attacks instead of recognition accuracy on benign datasets. Characterise chaotic behaviour quantitatively, with a Lyapunov exponent and a bifurcation diagram, rather than by phase-portrait resemblance. Audit

cryptographic implementations for timing and side channels before offering them for use. Scale the experiment to the device count the deployment implies. Release the artefact, and check that the link resolves, which in this volume is not a rhetorical flourish (§4.2).
None of these requires new apparatus. Each is a study the same authors could run on the same equipment, and each converts a claim the paper currently asserts into one it can support. This is the group we would recommend to the volume's authors first.

2. *The measurement the emerging technique needs.* Two problems are specific to §9 and they are complementary. The first is to benchmark typed tool contracts against function calling on a task suite with an error taxonomy, which is the experiment Zhakhalov [152] argues for and does not run. The second is sharper, and neither paper in the cluster raises it: to measure what constraining the action space costs in reasoning quality, given DOI-verified evidence that format restriction degrades reasoning and that the penalty falls hardest on instruction-tuned models [189, 190]. That item is the only one in table 7 carried by no author's future work; it comes from the discussion and the external literature alone, and its answer would determine whether the convergence reported in §9 is well founded.
3. *Longer-horizon problems, and the one that is not technical.* The remainder need more than a revision: multi-temporal predictive modelling of forest disturbance; runtime-adaptive agent behaviour in place of baked animation; validated Ukrainian-language resources and transformer baselines; hardware and software co-design with reportable conditions for neuromorphic systems; and measured learning effects for the educational artefacts, which requires a study design this volume's education papers do not yet have.

Table 7: Open problems surviving triangulation. A problem is listed only where at least two of three independent streams support it: the authors' own future work (A), what the discussion asked for (D), and the 2024–2026 literature (E). 23 candidate problems were grouped from 32 author future-work sentences, the recommendation, directive and extension asks, and the retrieved literature; 6 were carried by a single stream and are excluded. The grouping is editorial; the filter is not.

Open problem	Cluster	A	D	E	Horizon	What would count as progress
Validate perceptual thresholds with human subjects, including CVD [170]	C5	•	•	•	medium	discrimination measured on participants, against the predicted threshold
Extend change detection to multi-temporal and predictive modelling [166]	C2	•	•	•	medium	a multi-year series with disturbance agents distinguished
Adapt agent behaviour at runtime rather than replaying baked animation [191]	C5	•	•	•	medium	IK-driven posture responding to live input, with frame-time budget held
Validate on data the model has not seen, from another source [163]	C3	•	•	•	near	accuracy reported on an independent test set from a different collection
Evaluate liveness against presentation attacks, not recognition accuracy [163]	C4	•	•	•	near	APCER/BPCER on printed, replayed and mask attacks at a fixed threshold
Scale the experiment to the device counts the deployment implies [168]	C4	•	•	•	near	the protocol comparison repeated at hundreds of nodes
Report the sim-to-real discrepancy, not only the simulated result [167]	C2/C4	•	•	•	near	one platform calibrated against hardware, with the gap stated as a number
Benchmark typed tool contracts against function calling on a task suite [189, 163]	C1	•	•	•	near	success rate and error taxonomy over 20+ tasks against an RPC baseline
Estimate forecast uncertainty and robustness to lost observations [165]	C2	•	•	•	near	calibrated intervals, and error under simulated detection dropout
Co-design neuromorphic hardware and algorithms, and standardise the reporting [164]	C4	•		•	far	a latency-energy comparison under conditions another group can reproduce
Measure the learning effect of the educational artefacts, not their delivery	C6	•	•		medium	a pre/post study with a comparison group and a stated instrument
Build and validate Ukrainian-language resources and transformer baselines [164]	C1	•		•	medium	a validated lexicon, and a fine-tuned classifier as the comparison

continued on the next page

Table 7, continued

Open problem	Cluster	A	D	E	Horizon	What would count as progress
Treat wartime and resource constraints as a stated requirement, not context	C2/C4	•		•	medium	power, connectivity and operator-safety budgets declared with the evaluation
Release the artefact, and check that the link resolves [164]	all	•		•	near	a resolving DOI or repository for data and code, cited from the paper
Characterise chaotic systems quantitatively, not by phase-portrait resemblance	C4	•		•	near	Lyapunov exponent, bifurcation diagram, and basin analysis for hidden attractors
Audit cryptographic implementations for timing and side channels	C4	•		•	near	constant-time verification, and benchmarks against the reference implementation
Measure what constraining the action space costs in reasoning quality [189, 190]	C1		•	•	near	paired evaluation with the schema over the action only, and over the whole path

One item belongs to none of those groups and should not be lost among them. Treat wartime and resource constraints as a stated requirement, not as context. It is supported by an author's own future work and by a chair's recommendation in the discussion (not, it should be said, by anything in the retrieved literature) and what it asks for is concrete: power, connectivity and operator-safety budgets declared alongside the evaluation, so that a system's fitness for the conditions it was built for becomes something a reader can check. Twenty-two of this volume's 23 papers were written under those conditions (§4.1). At present the constraint shows up in motivation sections and almost never in evaluation sections, and closing that gap is within reach of any of them.

11. Limitations and threats to validity

1. *The authors are also the subject.* We chaired this workshop, edited this volume, wrote the prepared question sets discussed in §7.3, and authored two of the papers reviewed in §6. The safeguards described in this paper reduce the resulting exposures but do not remove them. Three are specific: two of the papers reviewed here are ours and were reviewed outside the submission system, as recorded in the conflict-of-interest statement; the comparison of prepared with delivered questions measures our own material, so the relation is causal instead of predictive; and the a-priori cluster partition tested in §5.1 is ours, which is why the run reported is the one declared in advance.
2. *No human coding pass.* The concept and facet coding was performed by five models from five vendors and by no human. It therefore measures model-independence and not human reliability, and it cannot establish that the codebook is correct, since five models might share a bias that no human would have and this design would not detect it. The statistic is reported as inter-model agreement throughout. A human double-coding pass is the largest single methodological improvement available to a repetition of this study.
3. *One dimension is not a measurement.* Response coding reached $\alpha = 0.37$ with 20 % unanimity, below any usable threshold, so §7.4 reports its distribution descriptively and builds nothing on it. Two facets are also weak: maturity at $\alpha = 0.41$, where the disagreement is substantive, and artefact availability at $\alpha = 0.10$, where it is prevalence skew and $\mathrm{AC}_1 = 0.63$ tells the more useful story. Unitisation reliability is not measured at all: the set of 59 asks was established manually and held fixed for every rater, so the agreement figures describe labelling and not segmentation.
4. *Measurement error in the transcript.* The discussion evidence is a single automatic transcript with no speaker labels and heavily garbled names. Attribution is by role only, no individual is named from it, and content words are never repaired. Eleven cues in one round render in the wrong alphabet (§7.1), four of 19 round boundaries are medium-confidence, and six answers are simply not recoverable. Where the recording is unclear the claim is dropped rather than reconstructed.

5. *Small n, one site, one session.* The study covers 23 papers, 19 discussion rounds, 59 asks and 20 papers with review scores. Ward clustering at $n = 23$ is defensible, whereas manifold learning and density-based clustering are not and were not used. The correlations reported here, $r = -0.12$ for novelty against evidence among them, are observations at this scale and are labelled as such. The floor-uptake figure in §7.3 rests on six asks.
6. *Language asymmetry.* Most presenters were working in a second language. Answers are coded on informational content alone and never on fluency, and no analysis in this paper treats hesitancy or grammar as evidence about a presenter.
7. *Methods designed and abandoned.* A technology-radar construction from citation contexts was designed and dropped: the corpus supplies 3 coupling edges over 253 pairs (§4.3), so the radar would have rendered as a ring of isolated points and implied a structure that is not present. Author keywords were dropped as a structural signal for the same reason and are reported descriptively only.
8. *Retrieval is one tool, one window.* The external positioning rests on 27 queries to a single literature-retrieval assistant, run once, in August 2026. Its recall is unknown; another tool, or the same tool in another month, would return a different shortlist. The shortlist rule is stated and its rejections are logged, which makes the selection auditable but not exhaustive. No reference in this paper is transcribed from a retrieval answer; every entry was rebuilt from its DOI, which is why the "Añón" placeholder that the assistant substitutes for unavailable author names never reached this bibliography.

12. Conclusion

The 23 papers of this volume do not form a thematically coherent collection, and the instruments conventionally used to demonstrate coherence would not have revealed the fact. Author keywords supply 154 tokens of which 146 are unique, and bibliographic coupling supplies 3 edges over 253 pairs. Drawn, both maps are fields of isolated points. The result is specific, not general: the profile of cited *venues* does carry structure, at 38 edges and a density of 0.150, so what fails is the appeal to shared handles and not bibliometric method as such. That distinction is what makes the finding useful to anyone reviewing a comparable volume.

The structure that does exist is recoverable semantically. A controlled vocabulary yields 33 shared concepts where author keywords yielded 6, and a stability-audited Ward partition of document embeddings agrees with the editors' independent grouping at ARI $= 0.34$. Where the two disagree, the disagreements are informative: the largest editorial cluster proves to be two groups, and the most similar pair of papers in the volume crosses an editorial boundary. Each stage is reported with its agreement statistics, including a facet whose α falls to 0.10 and a dimension too unreliable to report as a measurement at all.

Applying the same instruments to the recorded discussion yields a result the corpus alone does not contain. Two groups sharing no authorship, institution or reference converged within a single session on schema-constrained tool contracts as the action space for language-model agents, at cosine 0.650 against a median of 0.18 over 253 pairs. Neither of the empty maps above can detect the convergence, since the two papers share not one keyword. The claim is stated in falsifiable form in §9, positioned against DOI-verified 2024–2026 literature, and set against the strongest available counter-evidence, namely that constraining a model's output can degrade its reasoning, particularly where the model is instruction-tuned. The resulting position is conditional, and the experiment that would settle it is specified.

The discussion analysis also produced a result about the workshop as an event. 52 of the 59 asks came from a chair and 6 from the floor, across 3 rounds, at a workshop convened for young scientists; talks overran and question time absorbed the difference. These are measurements of a community's practice taken from its own artefacts.

The procedure applied here to one workshop can be applied to any workshop volume, and the

negative result that motivated it is likely to recur wherever a corpus is small, heterogeneous and written by researchers early in their careers. For this community the concrete output is the set of 17 open problems that survive triangulation in §10, nine of which are deliverable within a revision: validating on unseen data, reporting the sim-to-real gap, testing liveness against real attacks, and releasing artefacts with links that resolve.

Author contributions

Both authors chaired CS&SE@SW 2025 and edited this volume. **Conceptualization**, S.O.S. and A.M.S.; **methodology**, S.O.S.; **software** (the analysis pipeline, the coding harness and the generated tables and figures), S.O.S.; **validation** (the reliability computations, the partition-agreement analysis and the acceptance checks) S.O.S. and A.M.S.; **formal analysis**, S.O.S.; **investigation** (corpus reconstruction, transcript segmentation and coding), S.O.S. and A.M.S.; **data curation**, A.M.S.; **writing – original draft**, S.O.S.; **writing – review and editing**, A.M.S.; **visualization**, S.O.S.; **supervision** and project administration, A.M.S. Both authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data availability statement

The analysis pipeline, the codebook, the coding ledgers for all raters, the transcript segmentation with its cue fingerprints, the reliability computations, the external-positioning question set and all derived per-paper data are openly available at https://github.com/ssemerikov/cssesw2025-paper00-artefacts under CC BY 4.0. Every statistic in this paper is produced by a script in that repository, and running its generators reproduces every table and figure from the deposited data. The manuscript's own source is not deposited: the paper is published with this volume, and what the repository adds is the method behind it. Three items are deliberately withheld, for the reasons given in §1.3: the full automatic transcript, because it records identifiable early-career researchers speaking and only the derived role-labelled coding is needed to check the analysis; the mapping from reviewer pseudonyms to reviewer names; and the credentials for the retrieval tooling. The reviewed papers are cited individually and are open access in this volume.

Conflicts of interest

The authors declare a substantial and unavoidable conflict of interest. We are the programme chairs of the workshop under review and the editors of the volume it introduces, and every safeguard described in this paper is a response to that position.

Two of the papers reviewed here, Semerikov et al. [143] and Slobodianiuk and Semerikov [155], are authored by one of us. Neither passed through the submission system: each was reviewed externally by the other chair, who had no authorship interest in it. The prepared question sets discussed in §7.3 are also ours, so that analysis measures our own material and the relation between the prepared and the asked is causal instead of predictive. The a-priori cluster partition tested in §5.1 is ours as well, which is why the clustering run reported is the one declared in advance and not the one that agrees with it most closely. §11 sets out the exposures that remain, which these measures reduce but do not eliminate.

Acknowledgments

We thank the authors who submitted to CS&SE@SW 2025, the 31 members of the programme committee, and in particular the 17 of them who filed reviews. We thank the 6 additional reviewers who reviewed alongside them; between the two groups 92 reviews were filed, and the volume rests on them. We thank the participants whose questions are the evidence base for §7, including the 6 asked from the floor, which turned out to be the only part of the discussion no prepared script anticipated. The workshop was held online on 21 November 2025 under conditions of war and intermittent electricity supply; we thank everyone who presented and attended regardless.

We thank CEUR-WS.org for the editorial work that publishes this volume, and we are glad to count it among the workshop's supporters. Seven of the eight editions of this series appear in its proceedings, which is what made the series-level comparison in §3.1 possible at all: the earlier volumes are still there to be measured.

Declaration on Generative AI

This declaration follows the CEUR-WS policy on AI-assisting tools [192] and itemises every use, with the tool named and the function stated. The policy's own template applies throughout: the authors used the tools named below in order to perform the functions named below, and after using them the authors reviewed and edited the content as needed. The authors are fully responsible for the accuracy and integrity of this paper, and no AI tool is an author of it.

1. Language editing. Generative AI was used for grammar, spelling and sentence-level clarity, and for rephrasing individual sentences, on text written by the authors. All resulting text was reviewed and edited by the authors, who take responsibility for it.

2. Literature retrieval. Leapspace (Elsevier) was queried 26 times in August 2026 with the question set published in the artefact repository, and a deep-research run was made against the same tool; the 27 answers were used to locate relevant work. No text and no bibliographic record was transcribed from any answer. Every reference in this paper was rebuilt from its DOI by content negotiation, and 33 records carrying no resolvable DOI were excluded and logged. That safeguard was load-bearing rather than precautionary: the tool substitutes a placeholder string for author names it cannot resolve, which a copy-and-paste harvest would have carried into print.

3. Language models as the study's measuring instrument. This is a research method rather than a writing aid, and it is a substantial part of what the paper reports. Five models from five vendors (claude-opus-5 (Anthropic), minimax-m3 (MiniMax), gpt-oss:120b (OpenAI open weights), gemma4 (Google) and nemotron-3-ultra (NVIDIA)) independently coded all 23 papers against the published codebook, at temperature 0, from identical dossiers, on 2026-08-24. The 59 discussion asks were coded by the first author and by three of the same models over four seeds on the same date. Agreement is reported per code in §2.3 and §2.5 as *inter-model* agreement and never as inter-coder reliability, including one dimension whose reliability is too low to support a claim; §11 carries the absence of a human coding pass as a limitation. Every prompt, every rater's output and every reliability computation is deposited.

4. A language-model artefact as an object of study. Before the session the chairs generated 105 prepared chair questions with a language model, from the single instruction published in the artefact repository. §7.3 reports what became of that material, together with the confound that the chairs both wrote and used it. This is analysed material and not a use in preparing the manuscript.

5. What generative AI was not used for. No figure, diagram, chart or illustration in this paper was produced by a generative model. Every chart is a deterministic rendering of measured data, emitted as pgfplots or TikZ by the scripts in the artefact repository and reproducible from the deposited CSV files; the two schematic figures were drawn by hand in TikZ. The screenshot panels are photographic captures of the presentations themselves. Software was written with AI assistance, and every script was executed by the authors and its output inspected; the numbers in this paper come

from those executions rather than from any model's assertion.

A. Topic index of the volume

Of the 53 concepts in the codebook, 33 are carried by more than one paper and are indexed here with their ACM CCS anchors; 19 are carried by exactly one paper and are omitted, since a one-paper index entry is the paper's own keyword list by another name. This index is generated from the concept coding of §2.3, so it cannot drift from the analysis.

No SWEBOK-area index is provided. The design for this paper called for one, in the shape used by earlier prefaces in this series, but the project data contains no defensible mapping from these concepts to SWEBOK v4 knowledge areas, and constructing one by hand while presenting it as generated would be the practice §2 criticises. The substantive point survives the omission: this corpus does not decompose onto software-engineering knowledge areas — there is no SWEBOK area for airborne sensing or for colour perception — which is informative about what a young-scientists workshop in this field now contains.

B. Codebook and scoring rubrics

Codebook version 1.0, dated 2026-08-24. This is generated from the same file the raters were given, so what is printed here is what was applied.

B.1. Facets and their values

- **F1 — primary application domain**: D1 software engineering and tooling; D2 AI/ML methods and agents; D3 remote sensing and geospatial; D4 health, environment and industry; D5 security and cryptography; D6 signal and embedded systems; D7 graphics, VR and perception; D8 education, social science and disciplinary studies
- **F2 — primary technique**: T1 deep neural network, supervised; T2 classical ML or ensemble; T3 optimisation or metaheuristic; T4 LLM or generative model as a component; T5 analytical or mathematical modelling; T6 simulation; T7 systems and architecture engineering; T8 secondary study; T9 instrument, survey or expert method; T10 rule-based or lexicon-based method
- **F3 — data modality**: M1 image or video; M2 remote-sensing raster; M3 tabular or structured records; M4 text; M5 time series or 1-D signal; M6 3-D geometry or scene; M7 literature corpus; M8 none, not data-driven
- **F4 — evaluation strategy**: E1 held-out benchmark with standard metrics; E2 controlled comparison of alternatives on own data; E3 simulation-based measurement; E4 functional demonstration or prototype; E5 expert or user judgement; E6 synthesis of prior published results; E7 none reported
- **F5 — dataset provenance**: P1 named public benchmark; P2 public source, own extraction; P3 own collected or measured data; P4 synthetic or simulated; P5 published literature; P6 none
- **F6 — artefact availability**: A1 public repository link given; A2 artefact described but not linked; A3 no artefact
- **F7 — deployment target**: G1 server or cloud; G2 desktop application; G3 web application; G4 mobile or cross-platform; G5 embedded or microcontroller; G6 UAV or robotic platform; G7 XR headset; G8 none, analytical study
- **F8 — maturity, TRL-like**: 1 concept or position, no implementation; 2 analytical or simulated proof of principle; 3 prototype validated on offline data; 4 prototype validated in a realistic setting; 5 deployed or in use

B.2. Evidence-strength rubric

Additive, 0-6, published here so Figure 10 can be recomputed. Used only for the maturity x evidence-strength scatter, never as a quality judgement of a paper.

- **F4**: E1 = 2, E2 = 2, E3 = 1, E4 = 1, E5 = 1, E6 = 1, E7 = 0
- **F5**: P1 = 2, P2 = 1, P3 = 2, P4 = 1, P5 = 1, P6 = 0
- **F6**: A1 = 2, A2 = 1, A3 = 0

B.3. Adjudication rules

1. F2 on a comparative paper is the technique the paper RECOMMENDS or ADOPTS, not the set it compares. paper17 compares six models and recommends the random forest, so F2=T2; paper04 compares seven and recommends a GA+MLP hybrid, so F2=T3.
2. F1 is the application domain, which is deliberately NOT the same construct as the six clusters of section 5. Where they diverge (paper24: cluster C2 airborne sensing, F1 D5 security, because the application is counter-UAV airspace monitoring) the divergence is reported in section 5, not resolved by moving the paper.
3. F6=A1 requires a resolvable link to an artefact THE AUTHORS PRODUCED. Citing a third party's package or dataset is not A1 (paper15 cites colour-science on Zenodo; that is someone else's artefact, so paper15 is A3).
4. A concept is coded present only when the paper USES it, not when it is named in related work.

B.4. Concepts

- C01 **machine learning, general**. *Include*: any learned model used as the paper's method. *Exclude*: mere mention of ML in related work.
- C02 **deep neural networks**.
- C03 **convolutional networks**.
- C04 **transformer architectures**.
- C05 **recurrent and LSTM networks**.
- C06 **spiking neural networks**.
- C07 **ensemble methods**. *Include*: random forest, boosting, bagging.
- C08 **evolutionary computation**.
- C09 **mathematical optimisation**. *Include*: linear programming, greedy, Monte Carlo search.
- C10 **natural language processing**.
- C11 **large language models**.
- C12 **computer vision**.
- C13 **object detection and tracking**.
- C14 **image classification**.
- C15 **semantic segmentation**.
- C16 **speech and audio processing**.
- C17 **computer graphics and rendering**.
- C18 **virtual and augmented reality**.
- C19 **modeling and simulation**.
- C20 **dynamical systems and chaos**.
- C21 **graph databases**.
- C22 **query languages**.
- C23 **information retrieval and search**.
- C24 **data visualisation**.
- C25 **knowledge representation**.
- C26 **cryptography**.
- C27 **post-quantum cryptography**.
- C28 **authentication and biometrics**.
- C29 **software and system security**.
- C30 **wireless and radio systems**.
- C31 **IoT and edge computing**.
- C32 **network protocols**.
- C33 **real-time and low-latency systems**.

- C34 **embedded systems and microcontrollers**.
- C35 **energy efficiency**.
- C36 **performance measurement**.
- C37 **software architecture**.
- C38 **software quality and testing**.
- C39 **API and interface design**.
- C40 **tool support and automation**.
- C41 **UAV platform**. *Include*: the paper's contribution depends on an airborne uncrewed platform.
- C42 **remote sensing**. *Include*: satellite or airborne imagery as the data source.
- C43 **geospatial analysis**.
- C44 **human-computer interaction**.
- C45 **colour and visual perception**.
- C46 **computing and STEM education**.
- C47 **serious games and gamification**.
- C48 **health risk assessment**.
- C49 **environmental monitoring**.
- C50 **tool-calling contract**. *Include*: a typed or schema-validated interface that bounds what an autonomous model may execute.
- C51 **wartime and demining application**.
- C52 **secondary study methodology**. *Include*: the paper's object of study is other publications.
- C53 **computational social science**. *Include*: the paper's object is a social-science question pursued computationally.

References

[1] A. E. Kiv, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CEUR Workshop Proceedings 2832 (2020) 1–10. URL: https://ceur-ws.org/Vol-2832/paper00.pdf.

[2] A. E. Kiv, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, 4th Workshop for Young Scientists in Computer Science & Software Engineering, CEUR Workshop Proceedings 3077 (2021) i–xxxv. URL: https://ceur-ws.org/Vol-3077/intro.pdf.

[3] S. O. Semerikov, A. M. Striuk, Embracing emerging technologies: insights from the 6th Workshop for Young Scientists in Computer Science & Software Engineering, CEUR Workshop Proceedings 3662 (2023) 1–36. URL: https://ceur-ws.org/Vol-3662/paper00.pdf.

[4] S. O. Semerikov, A. M. Striuk, The evolving landscape of computer science and software engineering: trends, challenges, and future directions, CEUR Workshop Proceedings 3917 (2024) 1–46. URL: https://ceur-ws.org/Vol-3917/paper00.pdf.

[5] I. Fedorchenko, A. Oliinyk, A. Stepanenko, M. Mikhaylova, T. Fedoronchak, Y. Gofman, M. Khokhlov, Development of hybrid optimization and prediction approaches for personalized diet design, International Journal of Computing 25 (2026) 381–393. doi:10.47839/ijc.25.2.4664.

[6] K. Krippendorff, Reliability in content analysis.: Some common misconceptions and recommendations, Human Communication Research 30 (2004) 411–433. URL: http://dx.doi.org/10.1111/j.1468-2958.2004.tb00738.x. doi:10.1111/j.1468-2958.2004.tb00738.x.

[7] A. F. Hayes, K. Krippendorff, Answering the call for a standard reliability measure for coding data, Communication Methods and Measures 1 (2007) 77–89. URL: http://dx.doi.org/10.1080/19312450709336664. doi:10.1080/19312450709336664.

[8] B. Efron, Bootstrap methods: Another look at the jackknife, The Annals of Statistics 7 (1979). URL: http://dx.doi.org/10.1214/aos/1176344552. doi:10.1214/aos/1176344552.

[9] K. L. Gwet, Computing inter-rater reliability and its variance in the presence of high agreement, British Journal of Mathematical and Statistical Psychology 61 (2008) 29–48. URL: http://dx.doi.org/10.1348/000711006X126600. doi:10.1348/000711006x126600.

[10] A. R. Feinstein, D. V. Cicchetti, High agreement but low kappa: I. the problems of two paradoxes, Journal of Clinical Epidemiology 43 (1990) 543–549. URL: http://dx.doi.org/10.1016/0895-4356(90)90158-L. doi:10.1016/0895-4356(90)90158-l.
[11] F. Gilardi, M. Alizadeh, M. Kubli, ChatGPT outperforms crowd workers for text-annotation tasks, Proceedings of the National Academy of Sciences 120 (2023). URL: http://dx.doi.org/10.1073/pnas.2305016120. doi:10.1073/pnas.2305016120.
[12] C. Ziems, W. Held, O. Shaikh, J. Chen, Z. Zhang, D. Yang, Can large language models transform computational social science?, Computational Linguistics 50 (2024) 237–291. URL: http://dx.doi.org/10.1162/coli_a_00502. doi:10.1162/coli_a_00502.
[13] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, 2019, p. 3980–3990. URL: http://dx.doi.org/10.18653/v1/D19-1410. doi:10.18653/v1/d19-1410.
[14] K. Song, X. Tan, T. Qin, J. Lu, T.-Y. Liu, MPNet: Masked and permuted pre-training for language understanding, 2020. doi:10.48550/ARXIV.2004.09297.
[15] J. H. Ward, Hierarchical grouping to optimize an objective function, Journal of the American Statistical Association 58 (1963) 236–244. URL: http://dx.doi.org/10.1080/01621459.1963.10500845. doi:10.1080/01621459.1963.10500845.
[16] R. R. Sokal, F. J. Rohlf, The comparison of dendrograms by objective methods, TAXON 11 (1962) 33–40. URL: http://dx.doi.org/10.2307/1217208. doi:10.2307/1217208.
[17] L. Hubert, P. Arabie, Comparing partitions, Journal of Classification 2 (1985) 193–218. URL: http://dx.doi.org/10.1007/BF01908075. doi:10.1007/bf01908075.
[18] A. E. Kiv, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, First student workshop on computer science & software engineering, CEUR Workshop Proceedings 2292 (2018) 1–10. URL: https://ceur-ws.org/Vol-2292/paper00.pdf.
[19] A. E. Kiv, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, Second student workshop on computer science & software engineering, CEUR Workshop Proceedings 2546 (2019) 1–20. URL: https://ceur-ws.org/Vol-2546/paper00.pdf.
[20] O. Mamyrbayev, N. Morkun, A. Kupin, S. Semerikov, V. Solovieva, A. Striuk (Eds.), Proceedings of the 5th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2022), Kryvyi Rih, Ukraine, December 16, 2022, SCITEPRESS – Science and Technology Publications, Setúbal, 2022. doi:10.5220/0000166700003561.
[21] N. Cavus, Y. B. Mohammed, M. N. Yakubu, An Artificial Intelligence-Based Model for Prediction of Parameters Affecting Sustainable Growth of Mobile Banking Apps, Sustainability 13 (2021) 6206. doi:10.3390/su13116206.
[22] D. Ghimire, S. Charters, S. Gibbs, Scaling Agile Software Development Approach in Government Organization in New Zealand, in: Proceedings of the 3rd International Conference on Software Engineering and Information Management, ICSIM '20, Association for Computing Machinery, 2020, pp. 100–104. doi:10.1145/3378936.3378945.
[23] P. Rozehnal, R. Danel, Aspects of Distance Education in Combination with Home Offices, Information 12 (2021) 75. doi:10.3390/info12020075.
[24] R. Danel, B. Gajdzik, Integrated Subsystems of Materials and Information Flow for Continuous Manufacturing of Coal and Steel, Management Systems in Production Engineering 32 (2024) 174–184. doi:10.2478/mspe-2024-0017.
[25] R. Wolniak, B. Gajdzik, M. Grebski, R. Danel, W. W. Grebski, Business Models Used in Smart Cities—Theoretical Approach with Examples of Smart Cities, Smart Cities 7 (2024) 1626–1669. doi:10.3390/smartcities7040065.
[26] R. DANEL, B. GAJDZIK, L. ROPYAK, Trends in data processing control in continuous production systems, Scientific Papers of Silesian University of Technology. Organization and Management Series 2023 (2023) 93–105. doi:10.29119/1641-3466.2023.181.6.
[27] S. Day, E. Erturk, E-Learning objects in the cloud: SCORM compliance, creation and deployment

options, Knowledge Management and E-Learning 9 (2017) 449–467. doi:10.34105/j.kmel.2017.09.028.

[28] J. Mesarosova, K. Martinovicova, H. Fidlerova, H. H. Chovanova, D. Babcanova, J. Samakova, IMPROVING THE LEVEL OF PREDICTIVE MAINTENANCE MATURITY MATRIX IN INDUSTRIAL ENTERPRISE, Acta Logistica 9 (2022) 183–193. doi:10.22306/al.v9i2.292.

[29] H. Fidlerová, J. Prachař, P. Sakál, Application of Material Requirements Planning as Method for Enhancement of Production Logistics in Industrial Company, Applied Mechanics and Materials 474 (2014) 49–54. doi:10.4028/www.scientific.net/AMM.474.49.

[30] P. Bajdor, I. Pawełoszek, H. Fidlerova, Analysis and Assessment of Sustainable Entrepreneurship Practices in Polish Small and Medium Enterprises, Sustainability 13 (2021) 3595. doi:10.3390/su13073595.

[31] P. Richnák, H. Fidlerová, Impact and Potential of Sustainable Development Goals in Dimension of the Technological Revolution Industry 4.0 within the Analysis of Industrial Enterprises, Energies 15 (2022) 3697. doi:10.3390/en15103697.

[32] A. F. Bardamid, A. I. Belyaeva, V. N. Bondarenko, J. W. Davis, A. A. Galuza, I. E. Garkusha, A. A. Haasz, V. G. Konovalov, A. D. Kudlenko, M. Poon, I. V. Ryzhkov, S. I. Solodovchenko, A. F. Shtan, V. S. Voitsenya, K. L. Yakimov, Ion fluence and energy effects on the optical properties of SS mirrors bombarded by hydrogen ions, Physica Scripta T 103 (2002) 109–112. doi:10.1238/physica.topical.103a00109.

[33] O. Haluza, I. Kolenov, V. Sokolenko, V. Lytvynenko, I. Gruzdo, Determination of Porosity and Moisture Content of Granular Activated Carbon Using Sub-Thz Ellipsometry, Journal of Infrared, Millimeter, and Terahertz Waves 46 (2024). doi:10.1007/s10762-024-01029-1.

[34] O. B. Akhiiezer, O. A. Haluza, A. O. Savchenko, L. M. Lyubchyk, N. T. Protsay, M. O. Aslandukov, Methodology of project-based learning for training junior students in applied mathematics: general scheme of the educational process, Journal of Physics: Conference Series 2611 (2023) 012005. doi:10.1088/1742-6596/2611/1/012005.

[35] O. Haluza, O. Kostiuk, A. Nikulchenko, O. Akhiiezer, M. Aslandukov, TEMPLATE-BASED MODEL FOR SHORT-TERM FORECASTING OF THE NUMBER OF TRANSACTIONS IN RETAIL CLOTHING STORES, Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies (2022) 51–56. doi:10.20998/2079-0023.2022.01.08.

[36] M. Azarenkov, V. Lytvynenko, I. Kolenov, O. Haluza, A. Chupikov, V. Sokolenko, O. Roskoshna, M. Kanishcheva, V. Shatov, Thermographic Method of Activated Carbon Packing Quality Diagnostics in NPP Air Filters, East European Journal of Physics (2024) 398–404. doi:10.26565/2312-4334-2024-1-41.

[37] P. Hryhoruk, N. Khrushch, S. Grygoruk, Environmental safety assessment: a regional dimension, IOP Conference Series: Earth and Environmental Science 628 (2021) 012026. doi:10.1088/1755-1315/628/1/012026.

[38] C. Pribeanu, D. D. Iordache, From Usability to User Experience: Evaluating the Educational and Motivational Value of an Augmented Reality Learning Scenario, in: Affective, Interactive and Cognitive Methods for E-Learning Design, IGI Global, 2010, pp. 244–259. doi:10.4018/978-1-60566-940-3.ch013.

[39] S. O. Semerikov, I. O. Teplytskyi, V. N. Soloviev, V. A. Hamaniuk, N. S. Ponomareva, O. H. Kolgatin, L. S. Kolgatina, T. V. Byelyavtseva, S. M. Amelina, R. O. Tarasenko, Methodic quest: Reinventing the system, Journal of Physics: Conference Series 1840 (2021) 012036. doi:10.1088/1742-6596/1840/1/012036.

[40] A. V. Riabko, T. A. Vakaliuk, O. V. Zaika, R. P. Kukharchuk, V. V. Kontsedailo, Chatbot algorithm for solving physics problems, CEUR Workshop Proceedings 3553 (2023) 75–92. doi:10.55056/ceur-ws.org/vol-3553/paper5.pdf.

[41] S. Papadakis, A. E. Kiv, H. M. Kravtsov, V. V. Osadchyi, M. V. Marienko, O. P. Pinchuk, M. P. Shyshkina, O. M. Sokolyuk, I. S. Mintii, T. A. Vakaliuk, A. M. Striuk, S. O. Semerikov, Revolutionizing education: using computer simulation and cloud-based smart technology to facilitate successful open learning, CEUR Workshop Proceedings 3358 (2023) 1–18. doi:10.55056/ceur-

ws.org/vol-3358/paper00.pdf.
[42] S. Papadakis, A. E. Kiv, H. M. Kravtsov, V. V. Osadchyi, M. V. Marienko, O. P. Pinchuk, M. P. Shyshkina, O. M. Sokolyuk, I. S. Mintii, T. A. Vakaliuk, L. E. Azarova, L. S. Kolgatina, S. M. Amelina, N. P. Volkova, V. Y. Velychko, A. M. Striuk, S. O. Semerikov, Unlocking the power of synergy: the joint force of cloud technologies and augmented reality in education, CEUR Workshop Proceedings 3364 (2023) 1–23. doi:10.55056/ceur-ws.org/vol-3364/paper00.pdf.
[43] S. Papadakis, A. E. Kiv, H. M. Kravtsov, V. V. Osadchyi, M. V. Marienko, O. P. Pinchuk, M. P. Shyshkina, O. M. Sokolyuk, I. S. Mintii, T. A. Vakaliuk, L. E. Azarova, L. S. Kolgatina, S. M. Amelina, N. P. Volkova, V. Y. Velychko, A. M. Striuk, S. O. Semerikov, ACNS Conference on Cloud and Immersive Technologies in Education: Report, CTE Workshop Proceedings 10 (2023) 1–44. doi:10.55056/cte.544.
[44] O. V. Gayevska, H. M. Kravtsov, Approaches on the augmented reality application in Japanese language learning for future language teachers, Educational Technology Quarterly 2022 (2022) 105–114. doi:10.55056/etq.7.
[45] V. N. Kukharenko, B. I. Shunevych, H. M. Kravtsov, Distance course examination, Educational Technology Quarterly 2022 (2022) 1–19. doi:10.55056/etq.4.
[46] V. Kryzhanivskyy, V. Bushlya, O. Gutnichenko, R. M'Saoubi, J.-E. Ståhl, Computational and Experimental Inverse Problem Approach for Determination of Time Dependency of Heat Flux in Metal Cutting, Procedia CIRP 58 (2017) 122–127. doi:10.1016/j.procir.2017.03.204.
[47] S. C. Mukhopadhyay, N. K. Suryadevara, A. Nag, Wearable Sensors and Systems in the IoT, Sensors 21 (2021) 7880. doi:10.3390/s21237880.
[48] D. N. Kumar, Social and Sexual Exploitation of Women in Vijay Tendulkar's Sakharam Binder, The Creative Launcher 8 (2023) 113–119. doi:10.53032/tcl.2023.8.5.12.
[49] S. Parajuli, N. Kumar, Need and struggles of homeless people: A study based on Nepalese context, International Journal of Science and Research Archive 13 (2024) 289–295. doi:10.30574/ijsra.2024.13.2.2087.
[50] A. Kupin, Y. Sherstnov, Y. Osadchuk, O. Savytskyi, The main aspects of building cyber-physical systems for optimal regulation of reactive power flows in main stepdown substations of mining and processing plants, in: CEUR Workshop Proceedings, Vol-3790: Proceedings of the 12th International Conference Information Control Systems & Technologies (ICST 2024), ICST 2024, CEUR-WS.org, 2024, pp. 306–317. doi:10.55056/ceur-ws.org/vol-3790/paper27.pdf.
[51] S. S. Korniienko, P. V. Zahorodko, A. M. Striuk, A. I. Kupin, S. O. Semerikov, A systematic review of gamification in software engineering education, CEUR Workshop Proceedings 3844 (2024) 83–95. doi:10.55056/ceur-ws.org/vol-3844/paper04.pdf.
[52] A. Kupin, D. Zubov, A. Zhenishbekova, Smart Control of Temperature and Audio Monitoring Inside Beehive: IoT ESP8266 NodeMCU and Android Mobile Platforms, in: CEUR Workshop Proceedings, Vol-3513: Proceedings of the 11-th International Conference "Information Control Systems & Technologies", ICST 2023, CEUR-WS.org, 2023, pp. 239–249. doi:10.55056/ceur-ws.org/vol-3513/paper20.pdf.
[53] A. Aljarbouh, D. Zubov, A. Kupin, N. Shaidullaev, Traffic-sign Recognition for Visually Impaired Pedestrians in Kyrgyzstan: Two-keypoint SIFT/BRISK Descriptor with CameraX, in: Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume III: Intelligent Systems Workshop, ISW-CoLInS 2024, CoLInS, 2024. doi:10.31110/colins/2024-3/011.
[54] S. Semerikov, A. Kupin, I. Marynych, A. Makohonov, Collection Data and Visualization Preventive Maintenance Schedule, in: Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume III: Intelligent Systems Workshop, ISW-CoLInS 2024, CoLInS, 2024. doi:10.31110/colins/2024-3/012.
[55] S. Semerikov, D. Zubov, A. Kupin, M. Kosei, V. Holiver, Models and Technologies for Autoscaling Based on Machine Learning for Microservices Architecture, in: Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Systems. Volume I: Machine Learning Workshop, MLW-CoLInS 2024, CoLInS, 2024. doi:10.31110/colins/2024-1/022.

[56] A. Kupin, D. Zubov, M. Kosei, V. Holiver, Architecture of intelligent system for webservices scaling, in: CEUR Workshop Proceedings, Vol-3790: Proceedings of the 12th International Conference Information Control Systems & Technologies (ICST 2024), ICST 2024, CEUR-WS.org, 2024, pp. 273–284. doi:10.55056/ceur-ws.org/vol-3790/paper24.pdf.

[57] D. Zubov, A. Kupin, N. Shaidullaev, Indoor spatial cognition of the hearing/visually impaired: pentest-inspired WiFi heatmap on Android platform, in: CEUR Workshop Proceedings, Vol-3790: Proceedings of the 12th International Conference Information Control Systems & Technologies (ICST 2024), ICST 2024, CEUR-WS.org, 2024, pp. 285–294. doi:10.55056/ceur-ws.org/vol-3790/paper25.pdf.

[58] D. Zubov, A. Kupin, Avoiding Type I Errors in Image Processing with SIFT/BRISK-keypoints on Android Smartphones, in: CEUR Workshop Proceedings, Vol-4048: Proceedings of the 13-th International Conference on Information Control Systems & Technologies (ICST 2025), ICST 2025, CEUR-WS.org, 2025, pp. 161–171. doi:10.55056/ceur-ws.org/vol-4048/paper13.pdf.

[59] I. Pilkevych, O. Boychenko, N. Lobanchykova, T. Vakaliuk, S. Semerikov, Method of assessing the influence of personnel competence on institutional information security, CEUR Workshop Proceedings 2853 (2021) 266–275. doi:10.55056/ceur-ws.org/vol-2853/paper33.pdf.

[60] N. M. Lobanchykova, T. A. Vakaliuk, V. P. Korbut, S. M. Lobanchykov, Y. B. Krasnov, Features of designing systems for the formation of an internal microclimate of a high class of cleanliness of operating rooms of medical institutions, IOP Conference Series: Earth and Environmental Science 1415 (2024) 012124. doi:10.1088/1755-1315/1415/1/012124.

[61] O. Nonik, N. Lobanchykova, T. Vakaliuk, V. Osadchyi, O. Farrakhov, Approaches to Solving Proxy Performance Problems for HTTP and SOCKS5 Protocols for the Case of Multi-Port Passwordless Access, in: CEUR Workshop Proceedings, Vol-3654: Proceedings of the Workshop Cybersecurity Providing in Information and Telecommunication Systems (CPITS 2024), CPITS 2024, CEUR-WS.org, 2024, pp. 189–200. doi:10.55056/ceur-ws.org/vol-3654/paper16.pdf.

[62] T. A. Vakaliuk, O. M. Spirin, N. M. Lobanchykova, L. A. Martseva, I. V. Novitska, V. V. Kontsedailo, Features of distance learning of cloud technologies for the organization educational process in quarantine, Journal of Physics: Conference Series 1840 (2021) 012051. doi:10.1088/1742-6596/1840/1/012051.

[63] O. V. Dubolazov, A. G. Ushenko, Y. A. Ushenko, M. Y. Sakhnovskiy, P. M. Grygoryshyn, N. Pavlyukovich, O. V. Pavlyukovich, V. T. Bachynskiy, S. V. Pavlov, V. D. Mishalov, Z. Omiotek, O. Mamyrbaev, Laser Müller matrix diagnostics of changes in the optical anisotropy of biological tissues, in: Information Technology in Medical Diagnostics II, CRC Press, 2019, pp. 195–204. doi:10.1201/9780429057618-24.

[64] A. V. Morozov, T. A. Vakaliuk, I. A. Tolstoy, Y. O. Kubrak, M. G. Medvediev, Digitalization of thesis preparation life cycle: A case of Zhytomyr Polytechnic State University, CEUR Workshop Proceedings 3553 (2023) 142–154. doi:10.55056/ceur-ws.org/vol-3553/paper14.pdf.

[65] N. Lobanchykova, S. Kredentsar, I. Pilkevych, M. Medvediev, Information technology for mobile perimeter security systems creation, Journal of Physics: Conference Series 1840 (2021) 012022. doi:10.1088/1742-6596/1840/1/012022.

[66] H. Lee, B. Moon, A. H. Aghvami, Enhanced SIP for Reducing IMS Delay under WiFi-to-UMTS Handover Scenario, in: 2008 The Second International Conference on Next Generation Mobile Applications, Services, and Technologies, 2008, pp. 640–645. doi:10.1109/NGMAST.2008.63.

[67] J. Bae, B. Moon, Time synchronization with fast asynchronous diffusion in wireless sensor network, in: 2009 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2009, pp. 82–85. doi:10.1109/CYBERC.2009.5342158.

[68] H. Lee, J. H. Ji, B. H. Moon, A. Aye Maw, J. Lee, Synergistic Multi-UAV Path and Surveillance Planning Using APF-Enhanced PPO, in: AIAA SCITECH 2025 Forum, American Institute of Aeronautics and Astronautics, 2025. doi:10.2514/6.2025-1613.

[69] H. Lee, J. H. Ji, B. H. Moon, A. Aye Maw, J. Lee, Correction: Synergistic Multi-UAV Path and Surveillance Planning Using APF-Enhanced PPO, in: AIAA SCITECH 2025 Forum, American Institute of Aeronautics and Astronautics, 2025. doi:10.2514/6.2025-1613.c1.

[70] M. O'Grady, D. Langton, F. Salinari, P. Daly, G. O'Hare, Service design for climate-smart agriculture, Information Processing in Agriculture 8 (2021) 328–340. doi:10.1016/j.inpa.2020.07.003.

[71] V. P. Oleksiuk, J. A. Overko, O. M. Spirin, T. A. Vakaliuk, A secondary school's experience of a cloud-based learning environment deployment, CEUR Workshop Proceedings 3553 (2023) 93–109. doi:10.55056/ceur-ws.org/vol-3553/paper7.pdf.

[72] V. P. Oleksiuk, O. R. Oleksiuk, T. A. Vakaliuk, A model of application and learning of cloud technologies for future Computer Science teachers, CEUR Workshop Proceedings 3820 (2024) 82–101. doi:10.55056/ceur-ws.org/vol-3820/paper134.pdf.

[73] V. P. Oleksiuk, O. R. Oleksiuk, The practice of developing the academic cloud using the Proxmox VE platform, Educational Technology Quarterly 2021 (2021) 605–616. doi:10.55056/etq.36.

[74] N. Balyk, I. Grod, Y. Vasylenko, V. Oleksiuk, Y. Rogovchenko, Project-based learning in a computer modelling course, Journal of Physics: Conference Series 1840 (2021) 012032. doi:10.1088/1742-6596/1840/1/012032.

[75] V. P. Oleksiuk, O. R. Oleksiuk, Examining the potential of augmented reality in the study of Computer Science at school, Educational Technology Quarterly 2022 (2022) 307–327. doi:10.55056/etq.432.

[76] I. Banicescu, R. L. Carino, J. P. Pabico, M. Balasubramaniam, Overhead analysis of a dynamic load balancing library for cluster computing, in: 19th IEEE International Parallel and Distributed Processing Symposium, 2005. doi:10.1109/IPDPS.2005.320.

[77] S. Dhandayuthapani, I. Banicescu, R. L. Carino, E. Hansen, J. R. Pabico, M. Rashid, Automatic selection of loop scheduling algorithms using reinforcement learning, in: CLADE 2005. Proceedings Challenges of Large Applications in Distributed Environments, 2005., 2005, pp. 87–94. doi:10.1109/CLADE.2005.1520907.

[78] J. V. Pleto, D. Magcale-Macandog, J. Campang, Y. C. Cabillon, F. Natuel, N. J. Larida, J. Pabico, C. Predo, V. G. Paller, Assessment and Local Community Perception on the Water Quality of the Seven Crater Lakes of San Pablo City, Philippines, Journal of Environmental Science and Management 27 (2024) 50. doi:10.47125/jesam/2024_1/05.

[79] D. B. Magcale-Macandog, A. R. Salvacion, J. P. Pabico, K. N. Tingson, M. A. Reblora, J. D. Edrial, F. P. Lansigan, M. T. Zafaralla, Developing a Bayesian Model of Climate-Induced Lake Overturn in Talisay, Taal Lake, Springer Nature Singapore, 2023, pp. 101–119. doi:10.1007/978-981-99-0131-9_6.

[80] A. Borja, R. M. Amongo, J. Pabico, D. Suministrado, Design and Evaluation of a Machine Vision System and Mechatronic Drive for a Pneumatic Seed Meter for Corn, Philippine Journal of Agricultural and Biosystems Engineering 17 (2021) 49. doi:10.48196/017.01.2021.05.

[81] D. Dymek, M. Grabowski, G. Paliwoda-Pękosz, A PROPOSITION OF AN EMERGING TECHNOLOGIES EXPECTATIONS MODEL: AN EXAMPLE OF STUDENT ATTITUDES TOWARDS BLOCKCHAIN, Technological and Economic Development of Economy 28 (2021) 101–130. doi:10.3846/tede.2021.15702.

[82] S. A. MacGowan, F. Madeira, T. Britto-Borges, M. Warowny, A. Drozdetskiy, J. B. Procter, G. J. Barton, The Dundee Resource for Sequence Analysis and Structure Prediction, Protein Science 29 (2020) 277–297. doi:10.1002/pro.3783.

[83] V. Derbentsev, N. Datsenko, V. Babenko, O. Pushko, O. Pursky, Forecasting Cryptocurrency Prices Using Ensembles-Based Machine Learning Approach, in: 2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T), 2020, pp. 707–712. doi:10.1109/PICST51311.2020.9468090.

[84] A. O. Bielinskyi, I. Khvostina, A. Mamanazarov, A. Matviychuk, S. Semerikov, O. Serdyuk, V. Solovieva, V. N. Soloviev, Predictors of oil shocks. Econophysical approach in environmental science, IOP Conference Series: Earth and Environmental Science 628 (2021) 012019. doi:10.1088/1755-1315/628/1/012019.

[85] A. Adamov, S. Mehdiyev, E. Seyidzade, Good practice of data modeling and database design for UMIS. Course registration system implementation, in: 2014 IEEE 8th International Conference

on Application of Information and Communication Technologies (AICT), 2014, pp. 1–4. doi:10.1109/ICAICT.2014.7035949.

[86] E. S. Vagif, A. A. Zahir, Developing of the creative abilities of the pupils by the using the project on training method in the classes of the informatics in the general schools, in: 2011 5th International Conference on Application of Information and Communication Technologies (AICT), 2011, pp. 1–5. doi:10.1109/ICAICT.2011.6110955.

[87] A. Ganbayev, E. Seyidzade, Enhancing Customs Fraud Detection: A Comparative Study of Methods for Performance Measurement and Feature Improvement, in: 2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), IEEE, 2023, pp. 1–5. doi:10.1109/aict59525.2023.10313153.

[88] S. H. Lytvynova, S. O. Semerikov, A. M. Striuk, M. I. Striuk, L. S. Kolgatina, V. Y. Velychko, I. S. Mintii, O. O. Kalinichenko, S. M. Tukalo, AREdu 2021 - Immersive technology today, CEUR Workshop Proceedings 2898 (2021) 1–40. doi:10.55056/ceur-ws.org/vol-2898/paper00.pdf.

[89] P. V. Zahorodko, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, M. I. Striuk, H. M. Shalatska, Comparisons of performance between quantum-enhanced and classical machine learning algorithms on the IBM Quantum Experience, Journal of Physics: Conference Series 1840 (2021) 012021. doi:10.1088/1742-6596/1840/1/012021.

[90] S. Semerikov, S. Chukharev, S. Sakhno, A. Striuk, A. Iatsyshyn, S. Klimov, V. Osadchyi, T. Vakaliuk, P. Nechypurenko, O. Bondarenko, H. Danylchuk, Our sustainable pandemic future, E3S Web of Conferences 280 (2021) 00001. doi:10.1051/e3sconf/202128000001.

[91] S. O. Semerikov, S. M. Chukharev, S. I. Sakhno, A. M. Striuk, A. V. Iatsyshin, S. V. Klimov, V. V. Osadchyi, T. A. Vakaliuk, P. P. Nechypurenko, O. V. Bondarenko, H. B. Danylchuk, 3rd International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, IOP Conference Series: Earth and Environmental Science 1049 (2022) 011001. doi:10.1088/1755-1315/1049/1/011001.

[92] A. E. Kiv, V. N. Soloviev, S. O. Semerikov, A. M. Striuk, V. V. Osadchyi, T. A. Vakaliuk, P. P. Nechypurenko, O. V. Bondarenko, I. S. Mintii, S. L. Malchenko, XIV International Conference on Mathematics, Science and Technology Education, Journal of Physics: Conference Series 2288 (2022) 011001. doi:10.1088/1742-6596/2288/1/011001.

[93] A. M. Striuk, S. O. Semerikov, Professional competencies of future software engineers in the software design: teaching techniques, Journal of Physics: Conference Series 2288 (2022) 012012. doi:10.1088/1742-6596/2288/1/012012.

[94] S. O. Semerikov, A. M. Striuk, H. M. Shalatska, AI-assisted language education: critical review, Educational Dimension 4 (2021) 1–7. doi:10.31812/ed.623.

[95] S. O. Semerikov, M. P. Shyshkina, A. M. Striuk, M. I. Striuk, I. S. Mintii, O. O. Kalinichenko, L. S. Kolgatina, M. Y. Karpova, 8th Workshop on Cloud Technologies in Education: Report, CTE Workshop Proceedings 8 (2021) 1–69. doi:10.55056/cte.183.

[96] M. I. Striuk, A. M. Striuk, S. O. Semerikov, Mobility in the information society: a holistic model, Educational Technology Quarterly 2023 (2023) 277–301. doi:10.55056/etq.619.

[97] A. E. Kiv, S. O. Semerikov, A. M. Striuk, V. V. Osadchyi, T. A. Vakaliuk, P. P. Nechypurenko, O. V. Bondarenko, I. S. Mintii, S. L. Malchenko, XV International Conference on Mathematics, Science and Technology Education, Journal of Physics: Conference Series 2611 (2023) 011001. doi:10.1088/1742-6596/2611/1/011001.

[98] S. O. Semerikov, S. M. Chukharev, S. I. Sakhno, A. M. Striuk, A. V. Iatsyshin, S. V. Klimov, V. V. Osadchyi, T. A. Vakaliuk, P. P. Nechypurenko, O. V. Bondarenko, H. B. Danylchuk, V. O. Artemchuk, 4th International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, IOP Conference Series: Earth and Environmental Science 1254 (2023) 011001. doi:10.1088/1755-1315/1254/1/011001.

[99] A. M. Striuk, Problematic questions of software requirements engineering training, Educational Dimension 4 (2021) 90–101. doi:10.31812/educdim.v56i4.4441.

[100] A. E. Kiv, S. O. Semerikov, M. P. Shyshkina, A. M. Striuk, M. I. Striuk, Y. V. Yechkalo, I. S.

Mintii, P. P. Nechypurenko, O. O. Kalinichenko, L. S. Kolgatina, K. V. Vlasenko, S. M. Amelina, O. V. Semenikhina, 9th Workshop on Cloud Technologies in Education: Report, in: CEUR Workshop Proceedings, Vol-3085: Proceedings of the 9th Workshop on Cloud Technologies in Education (CTE 2021), CTE 2021, CEUR-WS.org, 2022. doi:10.55056/ceur-ws.org/vol-3085/paper00.pdf.

[101] S. O. Semerikov, T. A. Vakaliuk, I. S. Mintii, V. A. Hamaniuk, V. N. Soloviev, O. V. Bondarenko, P. P. Nechypurenko, S. V. Shokaliuk, N. V. Moiseienko, V. R. Ruban, Mask and Emotion: Computer Vision in the Age of COVID-19, in: Digital Humanities Workshop, DHW 2021, Association for Computing Machinery, 2022, pp. 103–124. doi:10.1145/3526242.3526263.

[102] I. S. Mintii, T. A. Vakaliuk, S. M. Ivanova, O. A. Chernysh, S. M. Hryshchenko, S. O. Semerikov, Current state and prospects of distance learning development in Ukraine, CEUR Workshop Proceedings 2898 (2021) 41–55. doi:10.55056/ceur-ws.org/vol-2898/paper01.pdf.

[103] D. S. Shepiliev, S. O. Semerikov, Y. V. Yechkalo, V. V. Tkachuk, O. M. Markova, Y. O. Modlo, I. S. Mintii, M. M. Mintii, T. V. Selivanova, N. K. Maksyshko, T. A. Vakaliuk, V. V. Osadchyi, R. O. Tarasenko, S. M. Amelina, A. E. Kiv, Development of career guidance quests using WebAR, Journal of Physics: Conference Series 1840 (2021) 012028. doi:10.1088/1742-6596/1840/1/012028.

[104] I. P. Varava, A. P. Bohinska, T. A. Vakaliuk, I. S. Mintii, Soft Skills in Software Engineering Technicians Education, Journal of Physics: Conference Series 1946 (2021) 012012. doi:10.1088/1742-6596/1946/1/012012.

[105] S. O. Semerikov, T. A. Vakaliuk, I. S. Mintii, V. A. Hamaniuk, V. N. Soloviev, O. V. Bondarenko, P. P. Nechypurenko, S. V. Shokaliuk, N. V. Moiseienko, D. S. Shepiliev, Immersive E-Learning Resources: Design Methods, in: Digital Humanities Workshop, DHW 2021, ACM, 2021, pp. 37–47. doi:10.1145/3526242.3526264.

[106] T. A. Vakaliuk, O. M. Spirin, O. V. Korotun, D. S. Antoniuk, M. O. Medvedieva, I. V. Novitska, The current level of competence of schoolteachers on how to use cloud technologies in the educational process during COVID-19, Educational Technology Quarterly 2022 (2022) 232–250. doi:10.55056/etq.32.

[107] Y. Pankiv, N. Kunanets, O. Artemenko, N. Veretennikova, R. Nebesnyi, Project of an Intelligent Recommender System for Parking Vehicles in Smart Cities, in: 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), volume 2, 2021, pp. 419–422. doi:10.1109/CSIT52700.2021.9648687.

[108] H. Lypak, N. Kunanets, N. Veretennikova, H. Matsiuk, T. Kramar, O. Duda, An Information System Project Using Augmented Reality for a Small Local History Museum, in: 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT), IEEE, 2023, pp. 1–4. doi:10.1109/csit61576.2023.10324194.

[109] R. Nebesnyi, N. Kunanets, R. Vaskiv, N. Veretennikova, Formation of an IT Project Team in the Context of PMBOK Requirements, in: 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT), IEEE, 2021, pp. 431–436. doi:10.1109/csit52700.2021.9648612.

[110] R. Vaskiv, N. Veretennikova, Integrated risk management model in distributed IT teams, Vìsnik Nacìonal'nogo unìversitetu "L'vìvs'ka polìtehnìka". Serìâ Ìnformacìjnì sistemi ta merežì 17 (2025) 214–225. doi:10.23939/sisn2025.17.214.

[111] F. Lin, A. Dewan, V. Voytenko, Open Interactive Algorithm Visualization, in: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), 2019, pp. 1–4. doi:10.1109/CCECE.2019.8861535.

[112] V. Voytenko, V. Vodichev, A. Kalinin, Comparative Analysis of Energy Performance of Induction Single-Motor and Multi-Motor Traction Electric Drive, in: 2021 IEEE 2nd KhPI Week on Advanced Technology (KhPIWeek), 2021, pp. 73–78. doi:10.1109/KhPIWeek53812.2021.9570063.

[113] A. De Renzis, M. Garriga, A. Flores, A. Cechich, A. Zunino, Case-based Reasoning for Web Service Discovery and Selection, Electronic Notes in Theoretical Computer Science 321 (2016) 89–112. doi:10.1016/j.entcs.2016.02.006.

[114] A. Zunino, M. Campo, Chronos: A multi-agent system for distributed automatic meeting

scheduling, Expert Systems with Applications 36 (2009) 7011–7018. doi:10.1016/j.eswa.2008.08.024.

[115] M. Hirsch, C. Mateos, A. Zunino, T. A. Majchrzak, T.-M. Grønli, H. Kaindl, A Task Execution Scheme for Dew Computing with State-of-the-Art Smartphones, Electronics 10 (2021) 2006. doi:10.3390/electronics10162006.

[116] P. Sanabria, T. F. Tapia, A. Neyem, J. I. Benedetto, M. Hirsch, C. Mateos, A. Zunino, New Heuristics for Scheduling and Distributing Jobs under Hybrid Dew Computing Environments, Wireless Communications and Mobile Computing 2021 (2021). doi:10.1155/2021/8899660.

[117] M. Hirsch, C. Mateos, A. Zunino, T. A. Majchrzak, T.-M. Grønli, H. Kaindl, A Simulation-based Performance Evaluation of Heuristics for Dew Computing, in: Proceedings of the 54th Hawaii International Conference on System Sciences, HICSS, Hawaii International Conference on System Sciences, 2021. doi:10.24251/hicss.2021.868.

[118] W. Kang, C. Anand, J. H. Park, Y. Choe, Evolutionary Factors Contributing to the Emergence of Prediction, Adaptive Behavior 33 (2025) 307–319. doi:10.1177/10597123251364746.

[119] F. Kazemi Vanhari, C. Anand, C. Welch, Analyzing Interview Questions via Bloom's Taxonomy to Enhance the Design Thinking Process, in: Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2025), Association for Computational Linguistics, 2025, pp. 582–593. doi:10.18653/v1/2025.bea-1.42.

[120] O. Owojaiye, A. R. Rajamani, A. Zavaleta Bernuy, C. K. Anand, ShapeCreator Mobile: Practicing Coding at Home in Low-Resource Environments, in: Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 2, ITiCSE 2025, ACM, 2025, pp. 719–720. doi:10.1145/3724389.3731253.

[121] S. Emdadi, S. Smith, C. K. Anand, Communicating StateCharts (CSC), in: Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 2, ITiCSE 2025, ACM, 2025, pp. 747–748. doi:10.1145/3724389.3731259.

[122] N. Osmani, L. M. Dutton, C. K. Anand, A Structure Editor for Elm, in: Proceedings of the 41st ACM/SIGAPP Symposium on Applied Computing, SAC '26, ACM, 2026, pp. 1330–1331. doi:10.1145/3748522.3780010.

[123] S. Kumar Bhuram, Secure Federated Learning for Automotive Supply Chains: A Hybrid Encryption Framework for Privacy-Preserving Demand Forecasting, Technix International Journal for Engineering Research 12 (2025). doi:10.56975/tijer.v12i6.158213.

[124] S. K. Bhuram, Edge-Cloud AI for Dynamic Pricing in Automotive Aftermarkets: A Federated Reinforcement Learning Approach for Multi-Tier Ecosystems, World Journal of Advanced Engineering Technology and Sciences 15 (2025) 126–135. doi:10.30574/wjaets.2025.15.3.0909.

[125] S. K. Bhuram, Federated Learning for Automotive Aftermarket Supply Chains: A Privacy-Preserving Framework for Predictive Maintenance Optimization, Global Journal of Engineering and Technology Advances 23 (2025) 216–223. doi:10.30574/gjeta.2025.23.3.0200.

[126] S. R. Gottimukkala, S. K. Bhuram, AI for automating financial consolidation and reporting, CRC Press, 2026, pp. 315–322. doi:10.1201/9781003647300-38.

[127] A. Dudnik, Y. Kravchenko, O. Trush, O. Leshchenko, N. Dakhno, V. Rakytskyi, Study of the Features of Ensuring Quality Indicators in Multiservice Networks of the Wi-Fi Standard, in: 2021 IEEE 3rd International Conference on Advanced Trends in Information Theory (ATIT), 2021, pp. 93–98. doi:10.1109/ATIT54053.2021.9678691.

[128] L. Kuzmych, D. Ornatskyi, V. Kvasnikov, A. Kuzmych, A. Dudnik, S. Kuzmych, Development of the Intelligent Instrument System for Measurement Parameters of the Stress - Strain State of Complex Structures, in: 2022 IEEE 4th International Conference on Advanced Trends in Information Theory (ATIT), 2022, pp. 120–124. doi:10.1109/ATIT58178.2022.10024222.

[129] I. Bakhov, Y. Rudenko, A. Dudnik, N. Dehtiarova, S. Petrenko, Problems of Teaching Future Teachers of Humanities the Basics of Fuzzy Logic and Ways to Overcome Them, International Journal of Early Childhood Special Education 13 (2021) 844–854. doi:10.9756/int-jecse/v13i2.211127.

[130] N. Dakhno, O. Barabash, H. Shevchenko, O. Leshchenko, A. Dudnik, Integro-differential Models with a K-symmetric Operator for Controlling Unmanned Aerial Vehicles Using a Improved Gradient Method, in: 2021 IEEE 6th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD), IEEE, 2021, pp. 61–65. doi:10.1109/apuavd53804.2021.9615431.

[131] N. Dakhno, O. Leshchenko, Y. Kravchenko, A. Dudnik, O. Trush, V. Khankishiev, Dynamic Model of the Spread of Viruses in a Computer Network Using Differential Equations, in: 2021 IEEE 3rd International Conference on Advanced Trends in Information Theory (ATIT), IEEE, 2021, pp. 111–115. doi:10.1109/atit54053.2021.9678822.

[132] A. Dudnik, Y. Kravchenko, O. Trush, O. Leshchenko, N. Dahno, Y. Ryabokin, Routing Method in Wireless IoT Sensor Networks, in: 2022 IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC), IEEE, 2022, pp. 1–6. doi:10.1109/saic57818.2022.9922998.

[133] P. Gyawali, A. Bautista-Hernández, A. H. Romero, Acceleration, agency, and foundations: ten influential arXiv directions in AI for materials science (2025), Journal of Physics: Materials 9 (2026) 021002. doi:10.1088/2515-7639/ae512f.

[134] M. Levchenko, M. Parkin, J. McEntyre, M. Harrison, Enabling preprint discovery, evaluation, and analysis with Europe PMC, PLOS ONE 19 (2024) e0303005. doi:10.1371/journal.pone.0303005.

[135] J. A. Teixeira da Silva, S. Moussa, P. Tsigaris, Standardizing Preprint Policies Are Needed to Clarify What Counts as a Prior Publication, Journal of Scholarly Publishing 56 (2025) 756–776. doi:10.3138/jsp-2024-0060.

[136] M. M. Kessler, Bibliographic coupling between scientific papers, American Documentation 14 (1963) 10–25. URL: http://dx.doi.org/10.1002/asi.5090140103. doi:10.1002/asi.5090140103.

[137] H. Small, Co-citation in the scientific literature: A new measure of the relationship between two documents, Journal of the American Society for Information Science 24 (1973) 265–269. URL: http://dx.doi.org/10.1002/asi.4630240406. doi:10.1002/asi.4630240406.

[138] M. Callon, J.-P. Courtial, W. A. Turner, S. Bauin, From translations to problematic networks: An introduction to co-word analysis, Social Science Information 22 (1983) 191–235. URL: http://dx.doi.org/10.1177/053901883022002003. doi:10.1177/053901883022002003.

[139] N. J. van Eck, L. Waltman, Software survey: VOSviewer, a computer program for bibliometric mapping, Scientometrics 84 (2009) 523–538. URL: http://dx.doi.org/10.1007/s11192-009-0146-3. doi:10.1007/s11192-009-0146-3.

[140] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, W. M. Lim, How to conduct a bibliometric analysis: An overview and guidelines, Journal of Business Research 133 (2021) 285–296. URL: http://dx.doi.org/10.1016/j.jbusres.2021.04.070. doi:10.1016/j.jbusres.2021.04.070.

[141] T. T. Quan, T. H. C. Nguyen, M. S. Nguyen, Noise reduction by using more reference signals in GSC beamformer, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 137–146. doi:10.55056/cssesw2025/paper11.

[142] H. Zemlianukhina, R. Voliansky, N. Volianska, Y. M. Arif, D. Halushko, Arduino-based chaotic generator with polar attractor, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 157–177. doi:10.55056/cssesw2025/paper34.

[143] S. O. Semerikov, P. P. Nechypurenko, T. A. Vakaliuk, I. S. Mintii, A. O. Kolhatin, Neuromorphic computing for ultra-low latency cognitive radio: a paradigm shift toward brain-inspired spectrum intelligence, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 208–221. doi:10.55056/cssesw2025/paper40.

[144] I. A. Hrytsai, V. P. Oleksiuk, Y. P. Vasylenko, Development of a user authentication module

based on computer vision and machine learning technologies, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 178–194. doi:10.55056/cssesw2025/paper14.

[145] A. V. Timenko, N. A. Kulykovska, S. S. Hrushko, V. V. Shkarupylo, Modeling and analyzing the performance of IoT systems using IoTSim-Edge: a comparative study of data transfer protocols, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 195–207. doi:10.55056/cssesw2025/paper23.

[146] D. S. Zolotopupov, L. V. Bulatetska, V. V. Bulatetskyi, The Kyber encryption algorithm and its implementation in .NET, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 147–156. doi:10.55056/cssesw2025/paper31.

[147] M. Y. Tiahunova, H. H. Kyrychek, A. Kazurova, Y. Knyryk, D. Tiahunov, Game engine impact on hardware resources, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 76–91. doi:10.55056/cssesw2025/paper08.

[148] Y. Y. Brilov, A. M. Striuk, The history of software engineering as the basis for computer games in the visual novel genre, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 129–136. doi:10.55056/cssesw2025/paper37.

[149] D. M. Bodnenko, O. V. Lokaziuk, S. P. Radchenko, V. V. Proshkin, S. D. Shymchyk, The development and application of VBA macros in the teaching of science and mathematics, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 92–107. doi:10.55056/cssesw2025/paper10.

[150] D. Markevych, S. Krepych, I. Spivak, R. Krepych, Approach to assessing the impact of motivation on employee performance in an inclusive environment, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 261–274. doi:10.55056/cssesw2025/paper18.

[151] N. Rudnichenko, V. Vychuzhanin, D. Shvedov, N. Shibaeva, I. Petrov, Intelligent system for analysis and assessment of population health harm risks by air pollution level, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 247–260. doi:10.55056/cssesw2025/paper17.

[152] V. Zhakhalov, A framework for efficient and secure LLM agency: a case for the GraphQL paradigm, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 222–231. doi:10.55056/cssesw2025/paper07.

[153] A. Baran, I. Spivak, S. Krepych, LLM and MCP driven e-commerce assistant over graph databases, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 335–349. doi:10.55056/cssesw2025/paper27.

[154] O. Lytvynenko, M. Kukhta, M. Yenin, Natural language processing for sociological research with a focus on preprocessing effects in sentiment analysis, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 350–360. doi:10.55056/cssesw2025/paper41.

[155] A. V. Slobodianiuk, S. O. Semerikov, Advanced deep learning methods for text generation in 2022-2024 literature: a systematic review, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 361–380. doi:10.55056/cssesw2025/paper42.

[156] O. I. Rolik, V. M. Smolij, N. V. Smolij, Basics of building a system for detection, accounting and visualization of explosive objects, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 232–246. doi:10.55056/cssesw2025/paper09.

[157] A. Kupin, Y. Osadchuk, M. Kosei, Comparative analysis of UAV simulation platforms: challenges, opportunities, and future directions, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 275–288. doi:10.55056/cssesw2025/paper19.

[158] O. Chaskovskyy, O. Sinkevych, S. Havryliuk, A. Nechepurenko, O. Pelyukh, M. Rozumovskyi, Random Forest and deep learning approaches for detecting forest cover changes, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 289–305. doi:10.55056/cssesw2025/paper22.

[159] I. V. Krak, M. O. Shurypa, M. O. Molchanova, O. V. Mazurets, O. V. Sobko, O. V. Barmak, Deep learning method for UAV detection and flight path forecasting in real-time, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 306–320. doi:10.55056/cssesw2025/paper24.

[160] O. V. Mazurets, O. O. Zalutska, M. O. Molchanova, V. I. Klimenko, L. V. Bukhantsova, O. V. Zakharkevich, Deep neural network-based classification of textile macrostructures by synthetic-to-natural fiber ratio, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 321–334. doi:10.55056/cssesw2025/paper26.

[161] A. Doroshuk, Colorimetric assessment of visual color difference between interface elements, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 108–118. doi:10.55056/cssesw2025/paper15.

[162] O. Sinkevych, Y. Sapsa, O. Storozhuk, Design and implementation of modular animated agents in VR using inverse kinematics and XR toolkit, in: S. O. Semerikov, A. M. Striuk (Eds.), Proceedings of the 8th Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2025), Kryvyi Rih, Ukraine, November 21, 2025, CEUR Workshop Proceedings, CEUR-WS.org, Aachen, 2026, pp. 119–128. doi:10.55056/cssesw2025/paper20.

[163] M. H. Assidiqi, D. Alghazzawi, S. Alarifi, L. Cheng, Benchmarking Reference-Free LLM Agent Robustness Under Schema, Policy, and Toolset Drift, IEEE Access 14 (2026) 79662–79672. doi:10.1109/access.2026.3696096.

[164] C. Kang, K. Li, A lifecycle-oriented survey of LLM-based agents: Workflow, graph augmentation,

and future directions, Neurocomputing 704 (2026) 134838. doi:10.1016/j.neucom.2026.134838.

[165] W. Wang, J. Fu, J. Song, K. Li, H. Qiao, J. Liu, H. Sun, X. Cao, Dist-Tracker: A Small Object-Aware Detector and Tracker for UAV Tracking, in: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2025, pp. 6603–6611. doi:10.1109/cvprw67362.2025.00657.

[166] V. Sineglazov, K. Lesohorskyi, Reference Point Planning for Copter Type UAVs for Orthophotomosaic Restoration in the Task of Explosive Object Detection, in: 2025 IEEE 8th International Conference on Methods and Systems of Navigation and Motion Control (MSNMC), IEEE, 2025, pp. 1–5. doi:10.1109/msnmc61130.2025.11399717.

[167] R. Tiwari, S. Khapre, A. Singh, Reinforcement learning in robotic systems : A review on sim-to-real transfer, Robotics and Autonomous Systems 198 (2026) 105327. doi:10.1016/j.robot.2025.105327.

[168] M. Kashyap, V. Sharma, A comparative analysis and implementation of CoAP and MQTT protocol for IoT communication, Life Cycle Reliability and Safety Engineering 14 (2025) 715–730. doi:10.1007/s41872-025-00324-7.

[169] M. Wijayanto, A. A. Razak, A. Muttaqin, Comparison of CoAP and HTTP Protocol Performance on ESP32 Microcontroller as an Air Quality Monitoring IoT Device, in: 2024 12th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS), IEEE, 2024, pp. 124–129. doi:10.1109/eeccis62037.2024.10840028.

[170] P. W. Limpisathian, S. C. Cox, Evolving CAM16-UCS for Modular Accessible Perceptibility and Simultaneous Contrast Detection Tool for Geovisualization, in: 2025 IEEE Visualization and Visual Analytics (VIS), IEEE, 2025, pp. 381–385. doi:10.1109/vis60296.2025.00082.

[171] N. Leburu, Trust-Aware Orchestration Architecture for LLM-Assisted Workflows in Multi-Tenant Enterprise Systems, IEEE Access 14 (2026) 97094–97117. doi:10.1109/access.2026.3706063.

[172] M. Altay, F. F. ŞİMŞEK, A Comparative Analysis of Random Forest and U-Net Deep Learning Approaches for Multi-Temporal Sentinel-2 Land Cover Classification, in: 2026 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), IEEE, 2026, pp. 234–238. doi:10.1109/m2garss67833.2026.11582333.

[173] N. Podoprigorova, F. Safonov, S. Podoprigorova, A. Tarasov, A. Shikhov, Recognition of Forest Damage from Sentinel-2 Satellite Images Using U-Net, RandomForest and XGBoost, in: 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), IEEE, 2024, pp. 1–6. doi:10.1109/reepe60449.2024.10479810.

[174] T. Boston, A. Van Dijk, R. Thackway, U-Net Convolutional Neural Network for Mapping Natural Vegetation and Forest Types from Landsat Imagery in Southeastern Australia, Journal of Imaging 10 (2024) 143. doi:10.3390/jimaging10060143.

[175] E. Kalantari, H. Gholami, H. Malakooti, A. R. Nafarzadegan, V. Moosavi, Machine learning for air quality index (AQI) forecasting: shallow learning or deep learning?, Environmental Science and Pollution Research 31 (2024) 62962–62982. doi:10.1007/s11356-024-35404-1.

[176] P. Verma, R. Verma, M. Mallet, S. Sisodiya, A. Zare, G. Dwivedi, Z. Ristovski, Assessment of human and meteorological influences on PM10 concentrations: Insights from machine learning algorithms, Atmospheric Pollution Research 15 (2024) 102123. doi:10.1016/j.apr.2024.102123.

[177] M. A. K. Jailani N, G. C. Mara, Feature Selection in Ozone Feature Space Impacts Performance in Gradient Boosting, Random Forest, Xgboost and Adaptive Boosting Regressors, in: 2024 International Conference on Current Trends in Advanced Computing (ICCTAC), IEEE, 2024, pp. 1–6. doi:10.1109/icctac61556.2024.10581262.

[178] C. Hu, B. Chen, X. Feng, F. Nian, J. Wang, T. Li, Swelling-ViT: Rethink Data-Efficient Vision Transformer from Locality, Springer Nature Singapore, 2024, pp. 32–46. doi:10.1007/978-981-97-8505-6_3.

[179] L. D. Solano-Santamaría, A. S. Solís-Manzano, E. M. Víquez-Mora, Empowering Gastrointestinal Diagnostics: A Comparison of ConvNeXt and ViT Models, in: 2025 IEEE VIII Congreso Interna-

cional en Inteligencia Ambiental, Ingenieria de Software y Salud Electronica y Movil (AmITIC), IEEE, 2025, pp. 1–8. doi:10.1109/amitic68284.2025.11214600.

[180] M. S. Ahmed, F. Ahmed, Parameter-Efficient Adaptation of Vision Transformers for Retinal OCT Classification, in: 2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence & Networking (QPAIN), IEEE, 2026, pp. 1–6. doi:10.1109/qpain69676.2026.11545861.

[181] A. Moghazy, A. Bendary, H. Eldeeb, A. D. Elbayoumy, Post-Quantum SSH: Achieving Production Performance Through Direct Algorithm Integration, in: 2025 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), IEEE, 2025, pp. 70–77. doi:10.1109/miucc66482.2025.11196771.

[182] R. Hou, Y. Gao, Y. Liu, J. Ming, Y. Zhou, MEML-KEM: A Memory-Efficient Implementation of ML-KEM for IoT Devices, Springer Nature Singapore, 2026, pp. 216–230. doi:10.1007/978-981-95-8417-8_16.

[183] U. Rana, A. Rai, U. Sangwan, Unity vs Unreal Engine: A Comparative Study, in: 2025 International Conference on Digital Innovations for Sustainable Solutions (ICDISS), IEEE, 2025, pp. 1–5. doi:10.1109/icdiss68238.2025.11320625.

[184] S. Berrezueta-Guzman, S. Wagner, Choosing the Right Engine in the Virtual Reality Landscape, IEEE Access 14 (2026) 13972–13985. doi:10.1109/access.2026.3657272.

[185] O. Sobchyshak, S. Berrezueta-Guzman, S. Wagner, Pushing the boundaries of immersion and storytelling: A technical review of Unreal Engine, Displays 91 (2026) 103268. doi:10.1016/j.displa.2025.103268.

[186] J. A. Hernandez, M. Colom, Reproducible research policies and software/data management in scientific computing journals: a survey, discussion, and perspectives, Frontiers in Computer Science 6 (2025). doi:10.3389/fcomp.2024.1491823.

[187] K. Riehl, A. Kouvelas, M. A. Makridis, Revisiting reproducibility in transportation simulation studies, European Transport Research Review 17 (2025). doi:10.1186/s12544-025-00718-9.

[188] D. Beyer, Evaluating Tools for Automatic Software Testing (Report on Test-Comp 2026), Springer Nature Switzerland, 2026, pp. 449–468. doi:10.1007/978-3-032-22774-4_23.

[189] Z. R. Tam, C.-K. Wu, Y.-L. Tsai, C.-Y. Lin, H.-y. Lee, Y.-N. Chen, Let Me Speak Freely? A Study On The Impact Of Format Restrictions On Large Language Model Performance., in: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track, Association for Computational Linguistics, 2024, pp. 1218–1236. doi:10.18653/v1/2024.emnlp-industry.91.

[190] M. and Schall, G. a. de Melo, The Hidden Cost of Structure: How Constrained Decoding Affects Language Model Performance, in: Natural Language Processing in the Generative AI Era, volume 2025 of *RANLP 2025*, Incoma Ltd. Shoumen, BULGARIA, 2025, pp. 1074–1084. doi:10.26615/978-954-452-098-4-124.

[191] H. Liu, X. Zhan, S. Huang, T.-J. Mu, Y. Shan, Programmable Motion Generation for Open-Set Motion Control Tasks, in: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2024, pp. 1399–1408. doi:10.1109/cvpr52733.2024.00139.

[192] CEUR-WS.org, Policy on the use of generative AI in CEUR-WS.org publications, 2024. URL: https://ceur-ws.org/GenAI/Policy.html.