GAIK Toolkit Demos

Umair Ali Khan & Akseli Leskinen

Haaga-Helia University of Applied Sciences

Demo link (**register for trying the tool**):
https://gaik-demo.2.rahtiapp.fi/
Demo recording:
https://www.youtube.com/watch?v=iJdfjEq1DbE

Co-funded by the European Union

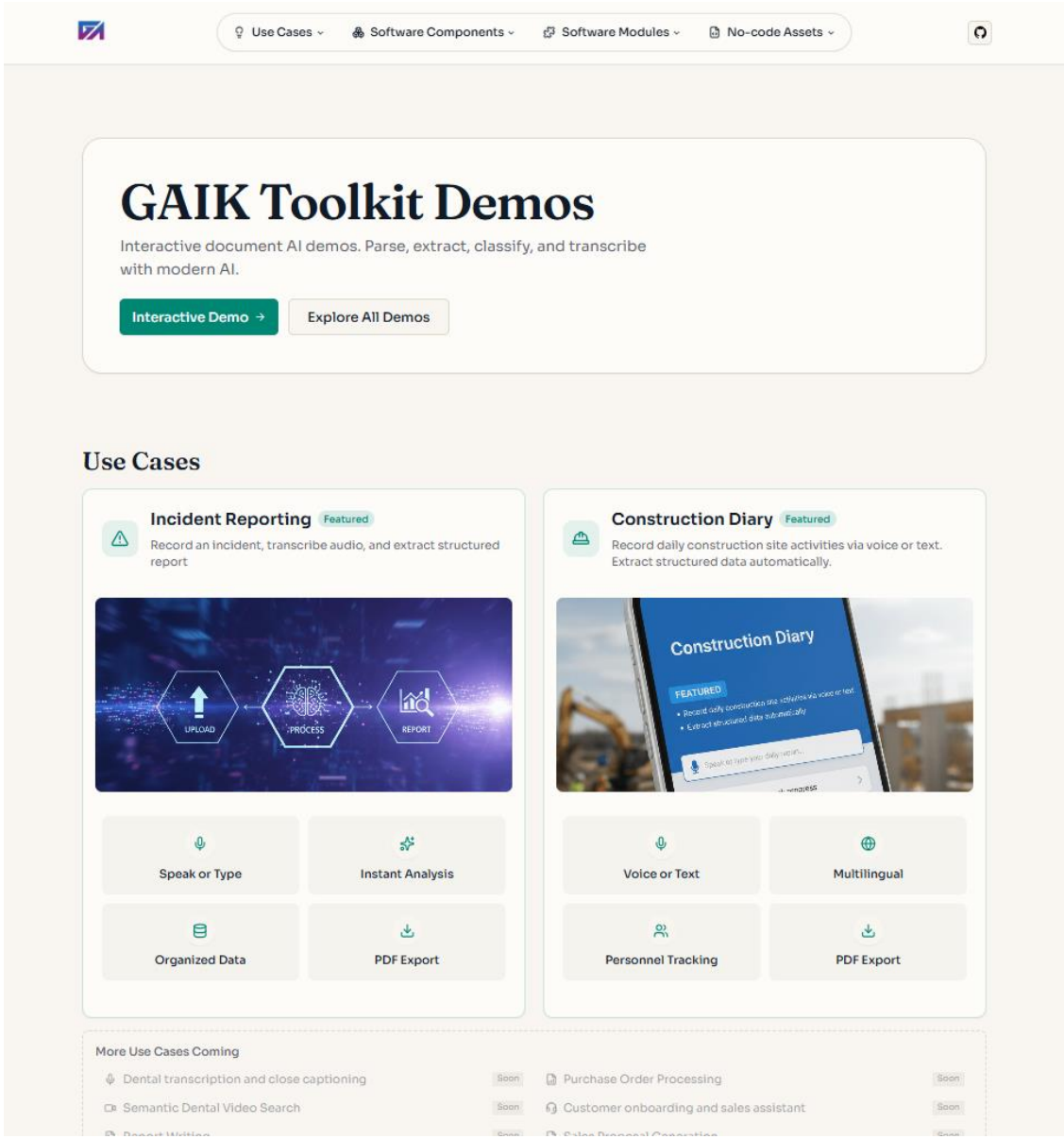
Haaga-Helia

UNIVERSITY OF HELSINKI

Tampere University

In collaboration with

FAIR Finnish AI Region | EDIH


Use Cases
Software Components
Software Modules
No-code Assets
GAIK Toolkit Demos
Interactive document AI demos. Parse, extract, classify, and transcribe with modern AI.
Interactive Demo →
Explore All Demos
Use Cases
Incident Reporting
Featured
Record an incident, transcribe audio, and extract structured report
UPLOAD
PROCESS
REPORT
Speak or Type
Instant Analysis
Organized Data
PDF Export
Construction Diary
Featured
Record daily construction site activities via voice or text. Extract structured data automatically.
Construction Diary
FEATURED
Voice or Text
Multilingual
Personnel Tracking
PDF Export
More Use Cases Coming
Dental transcription and close captioning
Soon
Purchase Order Processing
Soon
Semantic Dental Video Search
Soon
Customer onboarding and sales assistant
Soon

Examples of the GAIK components

- Business layer:
  - GenAI product canvas
  - GenAI product description: Incident reporting use case
- Strategic layer:
  - Value evaluation framework
  - Value evaluation model: Incident reporting use case
- Implementation layer:
  - Evaluation methods

GenAI product canvas (template)

BETA version

Name	The short, descriptive title of the AI solution or use case. It should clearly indicate what the assistant or system does.
Task	The main activity or job the AI solution is designed to support or automate. This describes *what is being done*, not how.
Knowledge processes	The knowledge-related activities supported by the AI system: knowledge capture (extracting information from inputs such as voice, images, or documents), knowledge access (retrieving relevant information from organizational systems and repositories), and knowledge synthesis (combining this information into a structured and usable output).
Business need	The problem or challenge the organization is trying to solve. This section explains *why* the solution is needed and what happens if the problem is not addressed.
Solution	A high-level description of how the AI-based system addresses the business need. It explains *what the solution does* and *how it works in practice*, without going into technical detail.
User/-s	The people who interact with or benefit from the solution. This can include employees, managers, customers, or other stakeholders.
Input	All information that the system needs in order to work. This includes data provided by users (for example, voice or images) as well as reference data, templates, or predefined lists used by the system.
Output	The result produced by the system after processing the input. This is usually a document, decision, recommendation, or structured data record.
Knowledge services	The specific ways in which the system supports knowledge work, such as guiding users, structuring information, retrieving relevant data, or ensuring consistency.
Expected benefits / value	The positive outcomes created by using the solution. These can include time savings, cost reduction, improved quality, better safety, or improved compliance.
Human-AI cooperation	A description of how responsibilities are shared between people and the AI system. This clarifies: Human: what users are responsible for (e.g. providing information, reviewing results, making final decisions) AI: what the system supports or automates (e.g. structuring data, suggesting categories, creating reports)

Co-funded by the European Union

GenAI product description: Incident reporting use case

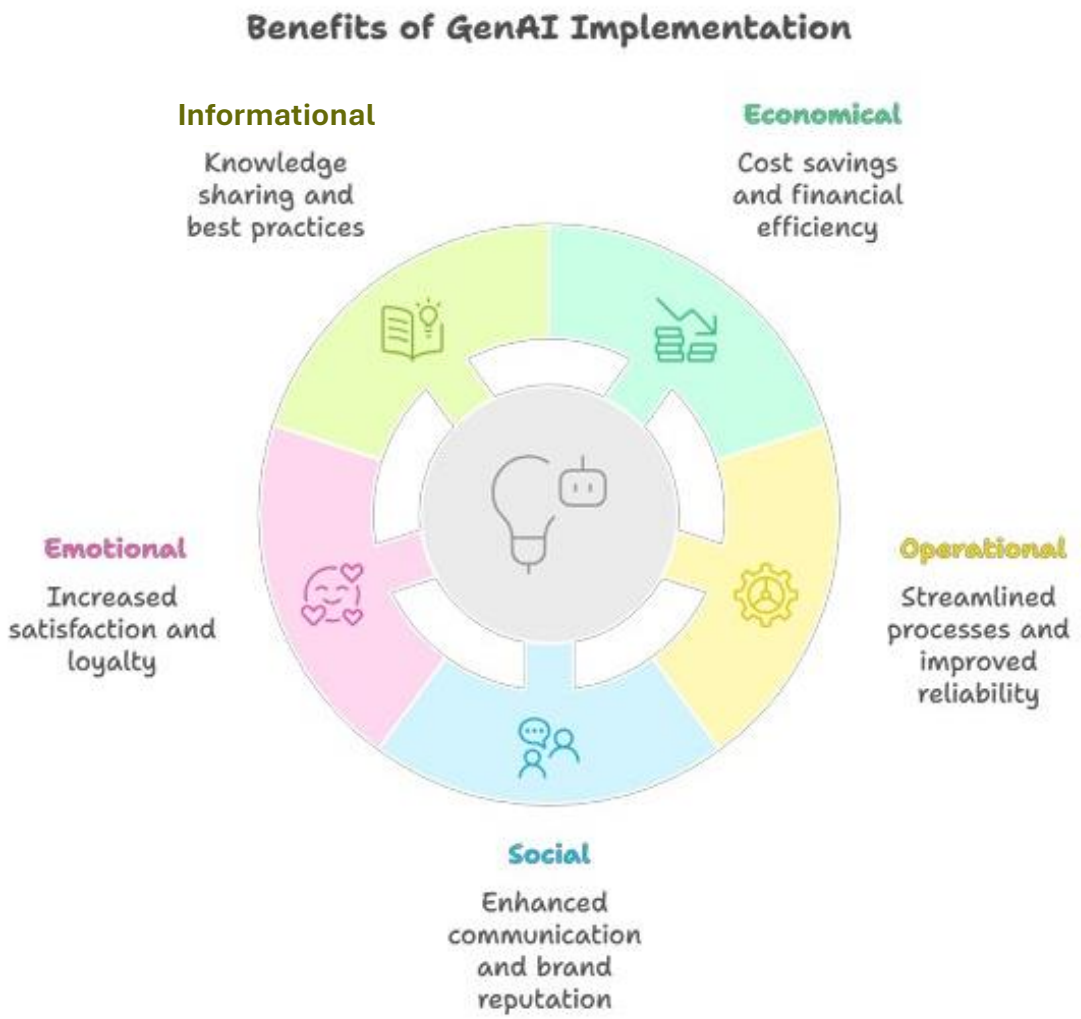
BETA version

Name	Incident reporting assistant
Task	Incident reporting
Knowledge processes	Knowledge Capture + Knowledge synthesis
Business need	Incidents such as broken equipment, water leaks, or spills occasionally occur on company premises and must be reported quickly so they can be resolved. Currently, employees must go to a computer and complete a web form in system X, which slows down the incident-reporting process.
Solution	The new AI-driven solution will enable employees to report incidents quickly through voice input on their mobile phones from different locations (including outdoors). The process will involve recording verbal descriptions of incidents, capturing images of hazards, and converting this information into a structured incident report that can be saved directly to the system X.
User/-s	Everyone in the company
Input	1. Voice input (Verbal description of an incident) 2. Images(photos), maybe with annotations 3. Template/-s for incident reports (list/-s of questions) 4. Reference data (list of equipment, list of facilities, list of incident types and severity levels)
Output	Incident report (filled in template)
Expected benefits / value	1. Faster Response Time 2. Increased Accuracy and Consistency 3. Improved Safety 4. Better Compliance and Documentation

Co-funded by the European Union

Value evaluation framework: Value dimensions

BETA version



Informational value – insight, understanding, decision quality

Economic value – time savings, productivity, ROI

Operational value – speed, flow, quality of work

Social value – sharing, collaboration, trust

Emotional value – agency, confidence, meaningful work

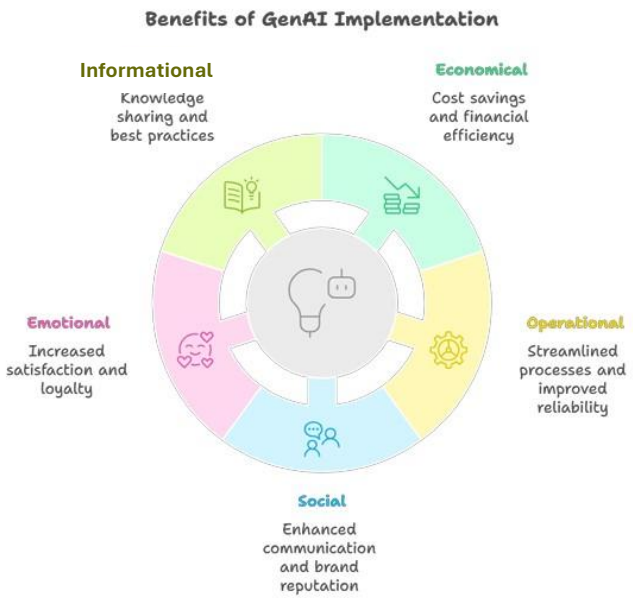
Value evaluation framework: metrics matrix

BETA version

Knowledge Process / Value Dimension	Capture	Access/ Retrieval	Synthesis
Informational Value	New knowledge documented	Time to relevant answer	Diversity of sources in decisions
Economic Value	Automated documentation rate	Time saved (€)	ROI of AI-supported decisions
Operational Value	Process learning captured	Reduction of rework	Reduced lead times
Social Value	Captured tacit knowledge	Transparency of sources	Cross-unit shared insights
Emotional Value	Recognition of individual expertise	Reduced frustration	Trust in AI-generated synthesis

Value evaluation model: Incident reporting use case

BETA version



⚙️ Functional Value (Primary)

- ⏱️ Faster reporting
- 🖐️ Less effort
- 📋 Complete, standardized reports
- 🌎 Accessible on-site

Outcome: More incidents reported, faster fixes

💰 Financial Value

- 📉 Lower admin effort
- 🛑 Accident cost avoidance
- 📈 Productivity gains

Outcome: Reduced operational costs

📊 Epistemic Value

- 🗂️ Better incident data
- 🔍 Improved insights
- 📈 Stronger analytics

Outcome: Smarter prevention decisions

❤️‍🔥 Emotional Value

- 🛡️ Higher confidence
- 🤝 Increased trust
- ⚡ Less reporting friction

Outcome: Employees feel safer and heard

🌱 Social Value

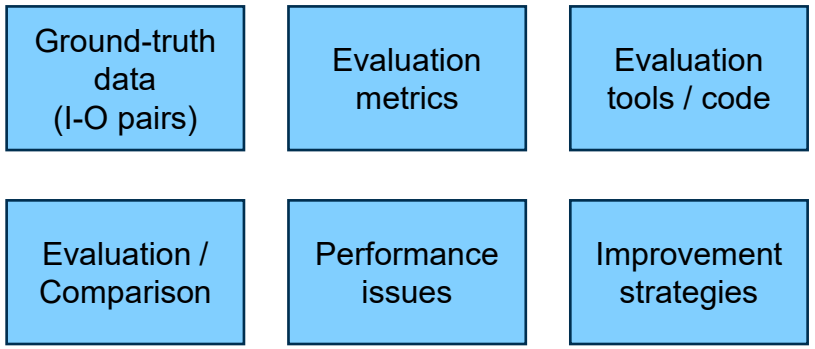
- 👥 Shared safety responsibility
- 🏅 Stronger safety culture
- 🌎 Corporate responsibility

Outcome: Better collaboration and reputation

Value Dimension	Relevance	Rationale
Functional	⭐⭐⭐⭐⭐ (Very High)	Core benefit: speed, ease, quality of reporting
Financial	⭐⭐⭐⭐☆ (High)	Cost avoidance and productivity gains
Epistemic	⭐⭐⭐☆ (Medium)	Better data and learning over time
Emotional	⭐⭐⭐☆ (Medium)	Trust, confidence, and motivation
Social	⭐⭐⭐☆ (Medium)	Safety culture and responsibility

Co-funded by the European Union

Incident reporting demo: Evaluation methods



Co-funded by the European Union

Evaluation methods: Strategies

Objective Evaluation

- Determining the output quality against certain metrics (e.g., accuracy)
- Compares system output with a correct reference
- Helps us fairly compare different systems
- Fast, consistent, and repeatable

Subjective Evaluation - Human Based

- Checks how good the output feels to a human
- Humans review for clarity, correctness, and readability
- Ensures the result makes sense and sounds natural

Evaluation methods: Objective Metrics

Transcription Evaluation

- **Word Error Rate (WER)** - How close the transcription is to the original audio
- **Missing Words (Deleted Rate):** Words that were skipped or left out
- **Extra Words (Added Rate):** Words added that weren't in the original
- **Substitution Error:** How often the model replaced a correct word with a wrong word.
- **Spelling Error:** Misspelled words

The lower the better.

Information Extraction

- **Exact Match Rate:** Certain fields (e.g., Name and Date) must match exactly.
- **Semantic Match Rate:** Certain fields (e.g., Event Description) should match approximately 70% in meaning.

Evaluation methods: Objective Metrics

Example metric description in the toolkit:

Understanding Word Error Rate (WER)

Definition

Word Error Rate (WER) is the standard metric for measuring speech-to-text accuracy. It quantifies the percentage of words that were incorrectly transcribed compared to a reference transcript.

Formula

```
WER = (S + D + I) / N × 100%
```

Where:

- **S (Substitutions)**: Words replaced with incorrect words
- **D (Deletions)**: Words omitted from the transcript
- **I (Insertions)**: Extra words added that weren't spoken
- **N**: Total number of words in the reference transcript

https://github.com/GAIK-project/gaik-toolkit/tree/main/implementation_layer/eval_methods/transcription_eval

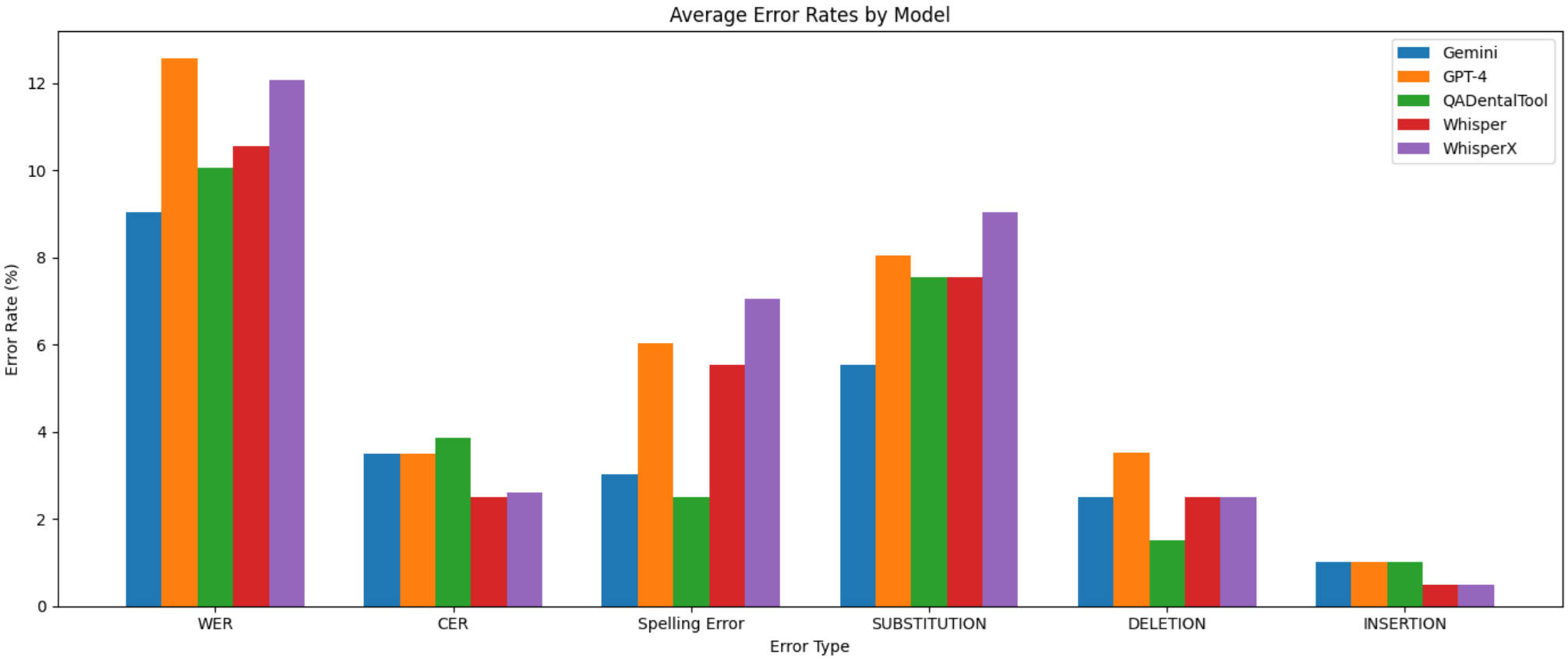
WER Interpretation Guidelines

WER Range	Assessment	Typical Applications
< 5%	Excellent	High-quality dictation, closed captions, medical transcription
5-10%	Very Good	Voice assistants in clean conditions, professional transcription
10-20%	Good	Meeting transcription, general-purpose STT
20-30%	Fair	Noisy environments, casual conversations
> 30%	Poor	Very challenging audio (heavy accents, background noise)

Source: What is WER in Speech-to-Text - Vatis Tech 2025

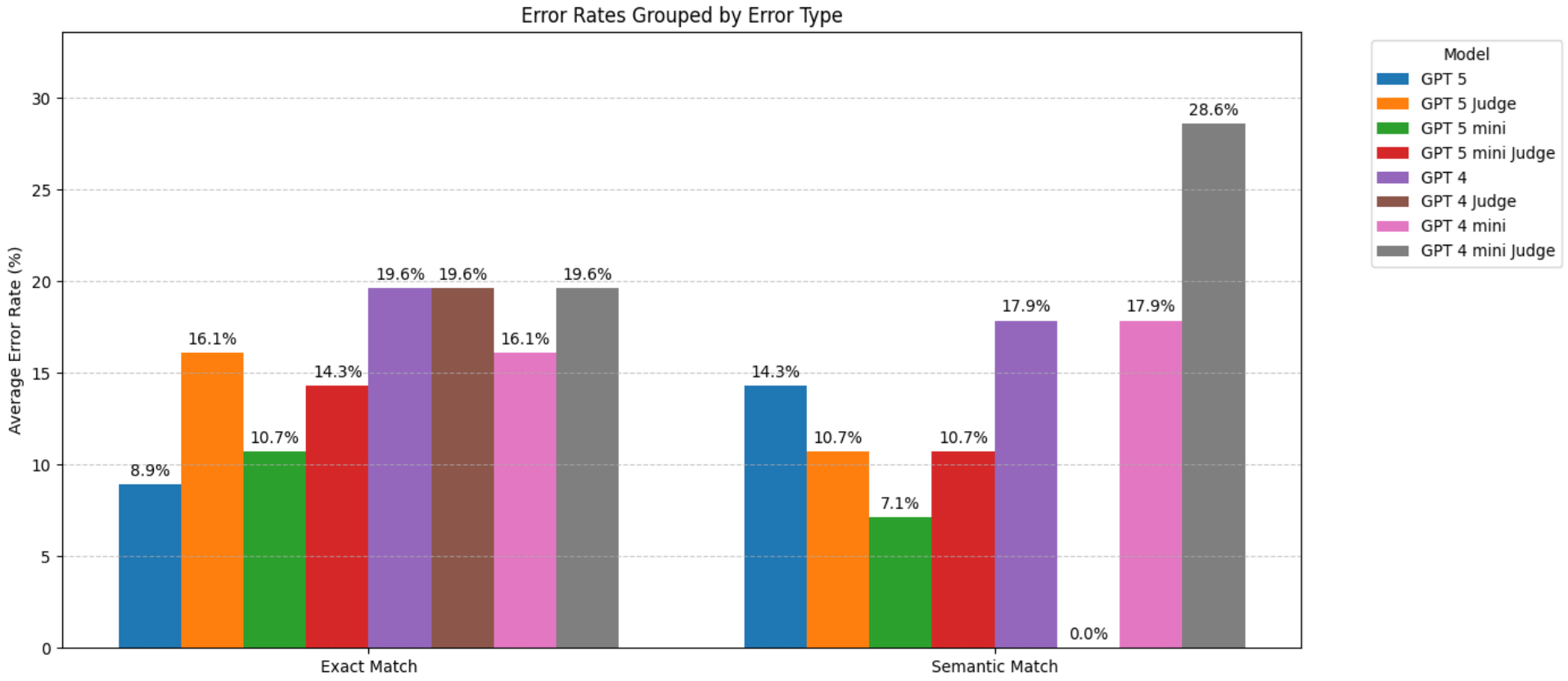
Evaluation methods: Performance comparisons

Transcription task



Evaluation methods: Performance comparisons

Information Extraction



AI performance issues & how to fix them

Transcription task

Error type	Description	Examples (from eval reports)	Fix vs live with (and fix stage)
Brands / proper nouns garbled	Brand/company names become phonetically similar nonsense tokens → high-impact substitutions	Straumann → Strauman, Dentsply Sirona → (split/garbled variants), Nobel Biocare → Nobel / Biocare distortions, Camlog Conelog → Camlog / Conelog variants, Implantona → implantoona/implanttoona	**Fix (high priority): Post-transcription**: entity normalization (standard brand list + fuzzy matching). If errors are extreme, a transcription-level prompt/glossary could help.
Names altered (identity drift)	Person names/surnames wrong (near-miss) → high user-visible errors	Martola/Martoon, Pallonen/Pallosen, Suojärvi/Suojärven	**Fix: Post-transcription**: spelling + dictionary-mapped variants only.
Digits vs Finnish number words	Reference uses words, but hypothesis uses digits (or vice versa) → full word errors	kahdenkymmenen → 20, yksitoista → 11 (Ari_Heiskanen_1_eval.txt); viisikymmentäkuusi → 56	**Fix: Post-transcription**: normalize to one policy (digits or words).
Decimals / spoken math tokenization	"X ja puoli", ranges, decimals become wrong single tokens or odd splits	... ja puoli → 375 (likely intended 37.5) (Ari_Heiskanen_1_eval.txt); kolmeviisi ↔ kolme viiva viisi	**Fix:** Mostly post-transcription with a Finnish number/decimal normalizer.
Function-word deletions (glue words missing)	Short words dropped: hurts WER but meaning often survives	frequent [D:ja], [D:että], [D:se] patterns	Often **live with it** if the meaning is intact. A post-transcription insertion of a small whitelist (että/ja/niin/…).
Unhelpful insertions (extra fillers/words)	Extra small words inserted → slight WER increase	insertions like ... [I] around sentence starts and discourse words	Usually **live with** unless the meaning changes. If needed: post-transcription filter repeated filler duplicates; avoid aggressive "helpful" rewriting.

AI performance issues & how to fix them

Transcription task

Error type	Description	Examples (from eval reports)	Fix vs live with (and fix stage)
Finnish morphology / case ending errors	Same lemma but wrong case/number/ending → substitutions	e.g., plural/singular or case changes like ...-llä ↔ ...-lla type drift	Fix: **Post-transcription** Finnish-aware correction (inflection only when very clear). Avoid meaning-changing edits.
Compound + hyphenation inconsistencies	Compounds split/merged; hyphenated clinical terms vary	peri-implantiitti ↔ periimplantiitti variants (Ajokortti_eval.txt); compound merging/splitting causing deletions	**Fix: Post-transcription** consistency normalization.
Loanword / English term distortion	English technical tokens misheard into unrelated Finnish words	lowdose → loudausohjelmia (Atte_Lajunen_1080p_eval.txt); minor loanword variants in endo/prosthetics context	**Fix: Post-transcription** bilingual domain lexicon + consistency.
Catastrophic omissions (chunk missing)	Large consecutive deletions → not recoverable safely from text	Atte Lajunen shows very high deletions; long [D:...] runs	Must fix at the **transcription level (fine-tuning)**. Should not be done at the post-transcript level
Evaluation artifacts (style vs content)	Changes that are “better text” but penalized by WER	hyphen/spacing differences (hands-on ↔ hands on) and similar tokenization shifts	**Fix: Evaluation-level** normalization (hyphen/number policy) + post-transcription consistent formatting.

Learning Resources

ARTICLES WITH CODE SHARING:

- Khan, U. A. (2026, January 2). *I built a Claude skill for knowledge extraction & report writing*. Data Science Collective. https://medium.com/data-science-collective/i-created-a-claude-skill-that-turns-piles-of-messy-documents-media-into-a-structured-report-19e9950f93b2
- Khan, U. A. (2026, January). *How to extract everything from documents using AI*. Data Science Collective. https://medium.com/data-science-collective/how-i-enhanced-doclings-image-interpretation-capabilities-641ce017bce5
- Khan, U. A. (2025, December 14). *Building reusable knowledge extraction AI workflows*. Data Science Collective. https://medium.com/data-science-collective/building-reusable-knowledge-extraction-ai-workflows-with-a-few-lines-of-code-a5aff93c0e0
- Khan, U. A. (2025, July 3). *Creating a knowledge extraction AI agent*. Data Science Collective. https://medium.com/data-science-collective/creating-a-knowledge-extraction-ai-agent-697e94f44afb

PEER-REVIEWED ARTICLES:

- Kudryavtsev, D., Khan, U. A., Remes J., Kauttonen J. (2025). Reuse and guidance for generative AI solution development and implementation: Knowledge management perspective. EDOC-CBI Forum 2025. 29th International Conference on Enterprise Design, Operations, and Computing & 27th International Conference on Business Informatics (CBI 2025), Lisbon, Portugal
- Kudryavtsev, D., Khan, U. A., Kauttonen, J., Kaski, T., Remes, J., Wuokko, A., Yangarber, R., Pivovarova, L., Wu, Y., Seppänen, M., Myllärniemi, J., & Sorri, K. (2025). Building a Generative AI toolkit for leveraging knowledge processes: The GAIK project report. 26th European Conference on Knowledge Management (ECKM 2025), Lahti, Finland. pp. 1317-1321.
- Khan, U. A., Kudryavtsev, D., Kauttonen, J., Joutsenniemi, A., Leskinen, A., Remes, J., Yangarber, R., Pivovarova, L., & Wu, Y. (2025). Evaluating generative AI technology choices and software frameworks for developing AI solutions in business. 26th European Conference on Knowledge Management (ECKM 2025), Lahti, Finland. pp. 1289-1294.
- Kudryavtsev D., Khan U. and Kauttonen J. (2024). Transforming Knowledge Management Using Generative AI: From Theory to Practice. In Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management – Volume 3: KMIS; ISBN 978-989-758-716-0, SciTePress, pages 362-370.

Thank you!

Questions?

Contacts:

UmairAli.Khan@haaga-helia.fi

akseli.leskinen@haaga-helia.fi