AI-BASED SOFTWARE TO INCREASE FACT CHECKING'S SPEED, SCALE AND IMPACT (ID14)

Introduction

An increasing number of companies employ artificial intelligence (AI) or advanced robotics in their workplaces. As part of EU-OSHA's research on advanced robotic and AI-based systems for the automation of tasks and occupational safety and health (OSH), 11 case studies and 5 short case studies were developed that focus on workplaces that use these technologies.

The objective of a case study is to investigate the practical implementation of advanced robotic and AI-based systems for the automation of physical and cognitive tasks in the workplace. This includes researching their impact on workers and related OSH dimensions, specifically, how OSH is managed in relation to such systems. This will help companies, policymakers and researchers gain a better understanding of the drivers, barriers and success factors for safe and healthy implementation of these systems.

To identify such case studies, several key informants at the EU and international levels, including workers' representatives and industry associations, were consulted. The participating companies then filled out a questionnaire, providing information about their company, describing the technology they use and addressing OSH-relevant topics regarding task automation. These results were then categorised within a taxonomy published in EU-OSHA's report 'Advanced robotics, artificial intelligence and the automation of tasks: definitions, uses, policies and strategies and Occupational Safety and Health'. Finally, each case study presents key takeaways, based on the experience of each company.

General company description

This short case study is based on a registered charity in England that employs close to 50 people. They were founded a decade ago and primarily focus on fact checking in the realm of media publications. In addition, **they develop and deploy AI-based tools for a fact-checking process**. Next to utilising their AI-based tools, they also further build scalable, robust, automated fact checking tools for newsrooms and fact checkers around the world. Since 2015, they have developed an AI-based software to help increase the speed, scale and impact of fact checking. They not only provide this software for others to use but also work in the field of fact checking themselves.

The non-profit organisation highlights their status as an independent and impartial entity, with the sole focus on identifying and countering false information. This purpose is rooted in their understanding and experience of how false information promotes hate, damages health and hurts democracy. They focus on fact checking claims made by politicians, public institutions and journalists, as well as viral content online. A vital part of their process is that once they have identified a false claim, they provide a corrected statement. Furthermore, they stress the importance of accountability by contacting the sources of the misinformation to either provide evidence of their statements or correct their mistakes.

Over the next three years, they intend to use machine learning to dramatically improve and scale fact checking. The company is working with international experts to develop new tools and to deploy and evaluate them. They **explicitly state their technology is not attempting to replace fact checkers but to empower them with the best tools available**.

Description of the system

The primary technology used by this company is natural language processing-based AI-algorithms. These perform a number of different tasks throughout the fact-checking process.

Underlying this technology is a recent Bidirectional Encoder Representations from Transformers (BERT) model. This is **a transformer-based machine learning technique for natural language processing (NLP)**.

Safety and health at work is everyone's concern. It's good for you. It's good for business.

1

The company fine-tuned this model with their own annotated data. BERT is a tool released by a major technology company that is pre-trained on a substantial language database covering over 100 languages. **By adding one's own annotated data, the model can be trained to perform in specific domains or for specific purposes.**

The system continuously collects and monitors incoming data from news sites, social media platforms, speeches from live TV, online news outlets and others. They continuously add more monitoring inputs from different countries to ensure coverage that is more complete. The input is then scanned by the technology to filter out statements with a need for fact checking. The non-profit also utilises functions to define specific parts of any given input text that makes claims over factual matters as the verifiable part of any sentence. **The non-profit focuses on statements made by politicians or journalists. The system then labels statements in need of further fact checking.** Text that does not contain any claims is not processed further. The fact-checking process itself can be very complex and involves active research into the claim's topic. Hence, this preselection reduces the volume of data that needs to be reviewed and subsequently fact checked by journalists.

Future development plans for the technology include training a BERT-style model to predict match/no-match for sentences and then add in entity analysis. Additionally, algorithms are supposed to be developed that can identify repeated claims that use different words to describe the same claim. These tools would further reduce journalists' workload.

Another task of this **AI-system is real-time checks of input data to help spot even more claims and further identify patterns of language to improve automated fact checking**. Given a sentence, the real-time analysis tool attempts to identify its **topic, trend, values, dates and location**. If these markers can be identified, it compares them with the corresponding data through the UK Office for National Statistics API. The knowledge base of topics and verbs is continuously expanding.

While the workers still perform essentially the same task as before the AI was implemented, there is a change in how they perform it. The AI preselects what information is led through to the individual worker for detailed fact checking. This saves them time and energy as well as limits their exposure to potentially harmful content.
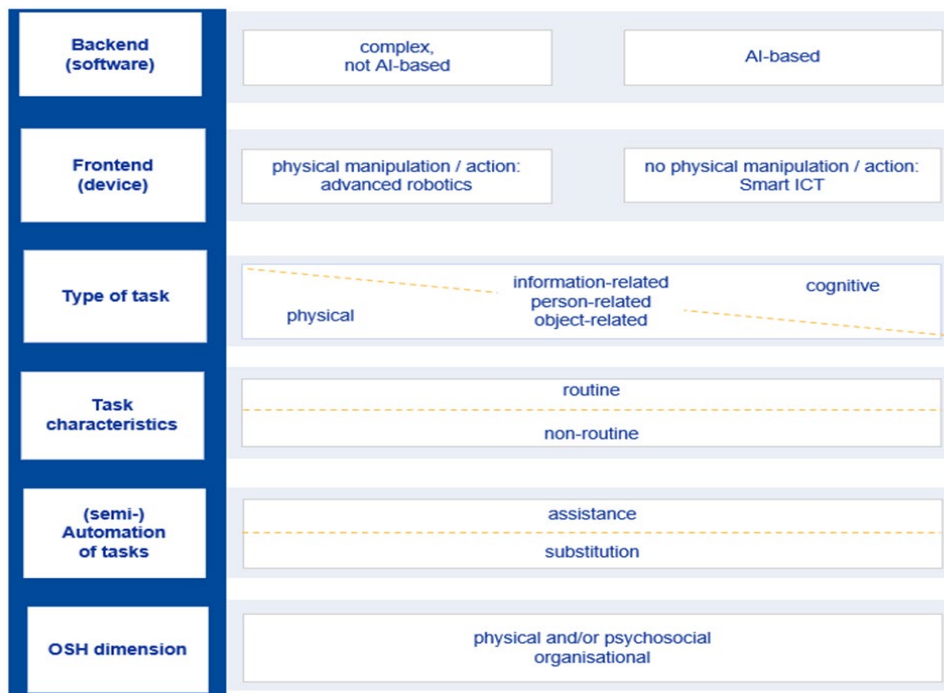
Taxonomy-based categorisation

To categorise different types of technology, a taxonomy specific for different important criteria of advanced robotic and AI-based systems was developed and published in the EU-OSHA report 'Advanced robotics, artificial intelligence and the automation of tasks: definitions, uses, policies and strategies and Occupational Safety and Health'.[1] This taxonomy includes the type of backend and frontend used and the type of task performed, as well as which category it falls under (information-related, person-related or object-related). It distinguishes between routine and non-routine task characteristics as well as the degree of automation in the forms of assistance or substitution. Finally, the taxonomy takes into account different OSH dimensions (physical, psychosocial and/or organisational) that are impacted by the technology.

---

[1] EU-OSHA – European Agency for Safety and Health at Work, *Advanced robotics, artificial intelligence and the automation of tasks: definitions, uses, policies and strategies and Occupational Safety and Health*, 2022. Available at: https://osha.europa.eu/en/publications/advanced-robotics-artificial-intelligence-and-automation-tasks-definitions-uses-policies-and-strategies-and-occupational-safety-and-health

Figure 1: Taxonomy for advanced robotics and AI-based systems for the automation of tasks

Backend (software)	complex, not AI-based	AI-based
Frontend (device)	physical manipulation / action: advanced robotics	no physical manipulation / action: Smart ICT
Type of task	physical	information-related, person-related, object-related — cognitive
Task characteristics	routine	
	non-routine	
(semi-) Automation of tasks	assistance	
	substitution	
OSH dimension	physical and/or psychosocial organisational	

The system automates a variety of **cognitive tasks**, based on an **NLP AI-based software**. This includes indexing and labelling text input and, in the future, content analysis like statement topic, value or location of origin. As described in detail in the previous section, the AI tools **assist** journalists in their work, rather than substitute them. Hence, the job content of the journalists remains unchanged. Only their work focus shifts more towards fact checking and away from text analysis. The OSH implications include mainly **psychosocial impacts**, reducing the amount of information a person has to process, by having the AI preselect and filter information. However, the company also takes **organisational efforts** to react to possible OSH implications, in the form of trainings and ethical audits.

OSH implications

A major psychosocial OSH risk for journalists and fact checkers is their exposure to harmful news content in order to fact check it. This includes reading articles depicting violence, hatred, conspiracy theories and so on. This content can be distressing to the reader. Using AI can potentially bring a different set of stories to the surface that would not be found by other tools. As the AI labels statements that have a heightened need for fact checking, it can **potentially increase the exposure of workers to harmful content**, depending on which statements the AI selects. However, the non-profit stated that they do not believe there are any systematic increases in exposure to harmful content resulting from the use of the AI technology. The company reports their awareness of existing OSH risks such as stress and mental hazards relating to journalists researching difficult topics, but the **AI tools do not increase such risks, in their experience**. To counteract these general negative effects, the company provides their workers with **peer support, regular breaks and personal flexibility** at their workplace. In addition, the company holds frequent open discussions to identify and avoid risks in their work. Relating to the AI tools, the company provides **training** for their workers to learn how to use the AI tools properly. This way, they can be used to benefit the workers by reducing the time they spend on identifying potential claims and increase the time they have to perform fact checking.

While concerns over deskilling caused by automating tasks is brought up in the discussion surrounding AI, this non-profit reports that their workers' core skills (journalism, fact checking and communication) are not impacted by the AI tools. The involved journalist also still performs the task that the AI performs, just to a different extent. Hence, they do not lose the underlying skill of performing the task manually. **The non-profit holds periodic 'ethical audits' of the tools to ensure there are no unintended or new negative social or ethical consequences.**

Key takeaways and transferability

The use and functionality of advanced robotic or AI-based systems can differ greatly from use case to use case. However, gathering information on similar use cases and transferring applicable insights about opportunities, risks and challenges, or other lessons learned, can help companies navigate the implementation process more efficiently and successfully, especially concerning OSH.

The company demonstrates heightened awareness regarding the ongoing discussion on **ethical implications** of AI-based systems. They have put countermeasures and checkpoints in the form of their company's internal ethical audits into place to address these early and continuously. While the used AI does not collect personal data, it does not mean that considerations of the ethical implications of using AI should be neglected. In this specific short case study, the implications of declaring statements by influential personnel (for example, people in high political offices) to be non-factual can have wide-ranging effects both on the person's and organisation's credibility, should the AI assessment be false and go unnoticed, or miss relevant statements. The introduction of **ethical audits** is one way to identify and address any rising concerns about the AI and the wider implications of using it.

In addition, the company fosters a workplace of **social support**, being aware of the OSH risks of journalism and fact checking (**for example, exposure to harmful or distressing content**), and the work with AI-based systems. The company shows awareness for potential **ethical risks** (involving an AI in the decision if statements are true or false) and takes preventive steps to identify and avoid them early on. This is indicative of an underlying work culture based on the awareness of the work the journalists are doing. Fostering a social surrounding that allows addressing ethical, functional or social concerns about the technology can be beneficial for the wellbeing of the workers involved.

This case study was commissioned by the European Agency for Safety and Health at Work (EU-OSHA). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect the views of EU-OSHA.