Journal of Artificial Intelligence Research 82 (2025) 2101-2124 Submitted 09/2024; published 04/2025

From Expert Systems to Generative Artificial Experts: A New Concept for Human-AI Collaboration in Knowledge Work

Konrad Sowa KOSOWA@KOZMINSKI.EDU.PL
Aleksandra Przegalinska APRZEGALINSKA@KOZMINSKI.EDU.PL

HumanRace Human-Machine Interaction Research Center, Kozminski University
Jagiellonska 57/59 St., 03-301 Warsaw, Poland

Abstract

This paper introduces Generative Artificial Experts (GAEs) - a concept of a new type of generative AI agents designed for human-AI collaboration in knowledge work. GAEs have specialized domain expertise, perform tasks within bounded autonomy, include a synthetic persona and possess multimodal generative AI capabilities, among other features. We provide a definition of GAEs which includes seven defining traits, offering a taxonomy which sets them apart from other generative AI systems. We use literature-review based conceptual analysis with abductive reasoning to propose the new concept that addresses identified limitations in existing systems. The paper explores the emergence of GAEs as a leap from expert systems. We name two enablers for GAEs - ongoing development of a research field of human-AI collaboration and growing capabilities of generative artificial intelligence systems. We discuss existing generative AI agents, noting that GAEs as such do not exist yet, but are starting to emerge. Due conceptual nature of this paper we do not explore the technical aspects of GAEs development. Instead, we use illustrative examples to present possible applications of GAEs and their potential role in the future of knowledge work.

1. Introduction

The field of generative artificial intelligence has witnessed substantial advancement in recent years. Thus far, access to this advanced technology was confined to a niche audience of software developers, researchers, and field experts. Recent developments, however, have made powerful tools publicly accessible. Public availability of generative AI tools and models, particularly large language models, has already notably altered professional workflows. Ongoing developments in agentic AI will likely enhance this trend (Acharya et al., 2025; Singh et al., 2024). These advancements have democratized the utilization of AI, bringing unprecedented capabilities to masses, however gaps in context-sensitive, specialized applications remain. Present generative models, despite their overall effectiveness, often fall short in delivering the in-depth domain understanding often necessary for specific tasks (Kumar et al., 2023; Mukherjee & Chang, 2023; Piller et al., 2024; Sharma et al., 2023).

Knowledge work is believed to be an area of work that will be the most impacted by a wider adoption of generative artificial intelligence. The impact of increased generative AI adoption in knowledge work is believed to be positive - reduced cognitive load, increased capacity for more complex tasks and enhanced learning process, are some of the benefits (Alavi & Westerman, 2023). Knowledge work provides possibly unlimited modalities of GAEs, thus the following conceptualization of generative artificial experts is based on that area. Furthermore, human-AI partnerships serve as enablers for enhancing knowledge work (Jarrahi, 2018).

This provides an opportunity to develop a new type of AI systems, specifically designed to meet the increasing demands for contextual intelligence in knowledge work. These systems could contribute to productivity more than existing generative AI systems. To address this gap, we develop and propose the concept of generative artificial experts (GAEs). Generative artificial experts (GAEs) are a new category of computer systems with conversational interface, based on generative artificial intelligence, and equipped with domain knowledge and tasked with solving problems together with people, in a collaborative manner. They act as semi-autonomous, highly specialized AI agents (Poole & Mackworth, 2023). GAEs have a narrow, but in-depth knowledge on a selected field, suitable for a given usage scenario. While they would include the general body of knowledge and comprehension of the world (alike other LLM-based systems), underlying models of GAEs would be trained on domain-specific training data to be able to solve specific problems in a way appropriate for that domain. We can imagine GAEs being applied in any knowledge work areas, from a local jurisdiction tax advisory, medical support, to help in running genetic sequencing in a lab. GAEs could support both experts and non-experts in knowledge work. Using GAEs would in many ways be similar to how LLM-based conversational systems are used. GAEs could assume specific personas, e.g. a seasoned university professor with hundreds of publications on a given topic or a successful entrepreneur. They would then adopt the communication style and overall attitude associated with said personas. Inclusion of the persona in GAEs design is justified by early results of studies showing that bots which resemble some human character traits are preferred by some users compared to those that do not. The persona's character traits are highly dependent on specific user preferences (Pal et al., 2023; Sowa et al., 2021). GAEs would be based on multi-modal generative artificial intelligence software, with a large-language model as a bedrock of the system. It means that they have capabilities including natural language processing, reasoning, learning, image processing and content generation. Thus, the interface for communication with users (people) is natural language - written or spoken. These systems would also be able to reason and make decisions, to a limited extent and with human oversight (bounded autonomy, as described later). It is also important to note that GAEs are not intended to replace human expertise in specialized domains, but to support it.

Current generative AI systems, while versatile, typically possess a broad and superficial knowledge base, making them prone to inaccuracies, or 'hallucinations,' when handling domain-specific tasks (Roychowdhury, 2023). Their reactive nature, reliant on user prompts, and limited capabilities for workflow integration often restrict their effectiveness. However, due to technological advancement in generative AI, new systems are closer to fulfilling the "emulating human expertise" (Jackson, 1999) promise than traditional expert systems were. This paper addresses a literature gap by introducing the novel concept of generative artificial experts, a system that uniquely combines multiple advanced traits to address the limitations observed in current generative AI systems. Our work is distinctive in its holistic approach, systematically unifying these traits into a framework that enhances AI's applicability in specialized knowledge work settings.

This paper employs a conceptual analysis, grounded in a review of literature across multiple domains including generative AI, expert systems, and human-AI collaboration. It utilizes abductive reasoning to propose a novel concept that addresses identified limitations in current generative AI systems.

In the first section we explain expert systems and their fundamental role for GAEs. Later we discuss generative artificial intelligence. Both concepts put together contribute to a leap from expert systems to generative artificial experts. In the second section we provide further theoretical grounding for GAEs - the growing body of knowledge about human-AI collaboration. The third section is focused on defining GAEs. We define the concept in detail, and present seven defining traits that differentiates it from other types of generative AI systems. We then show illustrative examples how GAEs could work, and we lay out practical instructions on how to develop GAEs using existing technologies. In section four we discuss the impact of GAEs, as well as concerns and challenges connected with their development.

2. From Expert Systems and Generative AI to Generative Artificial Experts

In this section we discuss expert systems and generative artificial intelligence, per their fundamental role for generative artificial experts. We briefly define expert systems and name their limitations. We highlight the increasing role and capabilities of generative AI and a possible leap to GAEs.

Generative artificial experts may be perceived as an evolution of expert systems, though far surpassing their capabilities and potential. Expert systems were the first concentrated attempt to encapsulate and replicate human expertise within a computational framework. In that sense, research on expert systems has laid the groundwork for the emergence of generative artificial experts. Expert systems were an important step in development of artificial intelligence, despite their low applicability at the time of their inception (1970s and later), as a result of limited computational capacity and other technical pitfalls that these systems had (Crevier, 1993; Partridge, 1987).

2.1 Expert Systems

Expert systems are artificial intelligence applications designed to solve complex problems by emulating the decision-making ability of a human expert. The essence of expert systems lies in their ability to synthesize vast amounts of information and reason using the gathered body of knowledge, offering specialized insights, much like a human expert would (Giarratano & Riley, 2005). Expert systems have found applications across various domains, such as data interpretation tasks (e.g. seismic analysis or semantic analysis in language processing), prediction tasks (e.g. weather forecasting), or medical tasks (e.g. medical diagnoses support) (Hayes-Roth et al., 1983). The foundations of expert systems were laid during the 1960s and 1970s, with pioneering projects such as DENDRAL (Lindsay et al., 1993), aimed at deducing chemical structures, and MYCIN (Shortliffe, 1974) , a system tailored for medical diagnoses. These attempts demonstrated the potential of rule-based systems to emulate human decision-making in niche domains. The subsequent decades, particularly the 1980s, witnessed an increase in research and commercial interest, resulting in the deployment of commercial expert systems, such as SID, written in LISP. In the late 20th and early 21st century expert systems were integrated with more advanced computational techniques. Harnessing the power of neural networks and rule based systems, the capabilities and adaptability of expert systems increased considerably (Yoon et al., 1994).

Expert systems consist of several interconnected components. The core of an expert system is the knowledge base, which stores specific information and rules. The inference engine leverages logical rules to deduce answers from the knowledge base. The explanation facility provides insight into the reasoning and decision-making process of the system. The knowledge acquisition facility streamlines the process of gathering expertise from human experts and translates it into a format that the expert system can utilize. Serving as the bridge between human users and the expert system, the user interface enables efficient interaction, and ensures that users can easily input queries and comprehend the system's responses (Giarratano & Riley, 2005; Waterman, 1985).

Traditional expert systems held several limitations. The knowledge acquisition bottleneck in expert systems was a significant challenge, as the process was very labor intensive - it required manual input and curation provided by human experts. Building, updating, and expanding the knowledge base was a time-consuming and costly process, making it difficult to keep the systems up-to-date and relevant (Chu & Hwang, 2008; Hubert & Dreyfus, 1997). Performance was also a challenge in early expert systems, often attributed to the use of tools that interpreted code expressions without prior compilation. As the knowledge base's size and complexity grew, so did the computational demands, making even simple tasks computationally expensive

(Partridge, 1987). Expert systems operated within a confined and static world of knowledge (often confined to a selected discipline of knowledge), lacking understanding of concepts and their interrelationships. This limitation made the systems unable to generate new insights or understand the connections between different elements of the knowledge base (Mccarthy, 1984).

2.2 Generative Artificial Intelligence

Generative artificial intelligence refers to a category of artificial intelligence systems that are designed to create new data, such as text, images, or audio, based on patterns learned from existing data (Goodfellow et al., 2016). The technology started growing in significance around 2017-2018 with development of first transformer-based systems, then gained even more traction in the following years (2019-2021), although the peak of applicability and development of generative AI is currently ongoing.

Some of the most prominent developments include large language models, such as OpenAI's GPT-1 through GPT-4, Meta AI's LLaMA, Gemini models from Google, or Claude models from Anthropic. Large-scale language models function based on the principle of statistical pattern recognition over extensive textual datasets. In the model training phase, they recognize linguistic structures and probabilistic dependencies within the textual data. Later, upon receiving a prompt, based on this body of knowledge, they produce human-like textual output. These models lack cognition - they do not possess semantic understanding of the text they get, similar to human cognition, but rather forecast subsequent textual elements based on previously learned patterns (Janik, 2023). On the other hand, knowledge bases that these systems were trained on were vast, thus the extent of information they can retrieve or produce is extensive too, in many ways convincingly similar to human domain experts (Z. Wang et al., 2022). Despite being young in development and application, LLMs are already considered as an immense value added for businesses, aiding in tasks such as translation, summarization, and question-answering (Brown et al., 2020).

Generative systems also provide significant capabilities in image generation. They are based on deep generative neural networks, utilizing diffusion models (Chang et al., 2023). Thanks to significant development of these methods, systems like Stable Diffusion, Midjourney or DALL-E can generate high quality images based on text descriptions (prompts).

In the software development sphere, systems like GitHub Copilot are able to understand and generate code in popular programming languages like Python, JavaScript, C and others (Yetistiren et al., 2022). There were also significant advancements in audio generation (Borsos et al., 2023), and there are ongoing attempts to generate video (Liu et al., 2024). Finally, multimodal generative AI (Suzuki & Matsuo, 2022), which combines different data modalities like text, images, and video, is an emerging area with immense application potential. By integrating various output types, these models will enable more complex and versatile applications.

Generative AI is foundational to GAEs. The multimodal capabilities of GAEs, including natural language processing and image or code generation, are direct extensions of the advancements in generative AI, allowing GAEs to interact and respond in ways that are most effective for the task at hand.

2.3 From Expert Systems to Generative Artificial Experts

Generative artificial experts are a technological leap from expert systems. This progression is rooted in the shared goal of emulating human expertise, with advancements in generative artificial intelligence being the cause for the significant technological leap.

The term 'experts' in 'generative artificial experts' directly references and extends the concept of 'experts' in 'expert systems.' In transitioning to generative artificial experts, we retain the notion of specialized proficiency but introduce broader capabilities. This evolution reflects not only advancements in AI technology but also a shift towards systems that are capable of learning, adapting, and, most of all, generating new insights beyond static rule-based reasoning. Therefore, while the name echoes the heritage of expert systems, it also marks a departure towards more sophisticated systems, aiming to support human expertise rather than merely replicate it.

When defining the 'expertise' of generative artificial experts, we follow the well-grounded works on expert systems. According to Giarratano and Riley (2005), expert systems are defined by their use of specialized knowledge to solve problems at the level of a human expert, and "an expert is a person who has expertise in a certain area". First expert systems were based solely on expert knowledge, however that loosened with time with many expert systems being based on explicit, codified knowledge. Giarratano and Riley propose a hierarchy, where true expertise - rare and specific to one domain, is higher than generally available knowledge, however expert systems should include both to function well. Authors add that "Expert systems, like human experts, are generally designed to be experts in one problem domain." and that "the expert system reasons or makes inferences in the same way that a human expert would reason or infer the solution of a problem". This understanding of the concept of an 'expert' directly translates to further work on generative artificial experts.

Expert systems, while groundbreaking, were fundamentally limited by their rigid rule-based frameworks, which restricted their ability to adapt to new or unforeseen scenarios, often resulting in error-prone performance in complex environments. These systems were isolated from the continuous feedback loops typical in dynamic settings, lacking the capability to evolve with the changing landscapes of their application areas, which rendered them obsolete as soon as the external parameters shifted. Expert systems also faced significant scalability issues. Their knowledge bases could not expand autonomously, requiring intensive manual intervention to learn or update, which starkly contrasts with the self-improving capabilities of modern generative AI. Moreover, the absence of collaborative and creative capacities in expert systems severely limited their usability in interdisciplinary and creative problem-solving tasks, areas where GAEs excel by design, leveraging continuous interaction and generative capabilities.

Expert systems laid the groundwork for generative artificial experts by demonstrating that AI can emulate human decision-making in specialized domains. By encapsulating expert knowledge into a series of rule-based operations, these systems could offer consultation and recommendations like a human expert. However, this emulation was often limited to specific, narrowly defined problem spaces, lacking the ability to learn from new data or adapt to evolving contexts. GAEs leverage advancements in generative AI, giving them the ability to process and generate diverse data forms, fulfilling more complex roles within their domain of expertise. Additionally, generative AI based systems can bring an element of creativity to problem-solving - they can produce novel solutions, ideas, and insights (Rick et al., 2023). They can manage larger scale tasks thanks to novel computational methods and higher processing power available. Furthermore, GAEs are not just isolated decision-makers but collaborative tools enhancing human capabilities.

3. Human-AI Collaboration as an Enabler for GAEs

In this section we explain how the ongoing development of the concept of Human-AI Collaboration provides the grounds for emergence of generative artificial experts. Prior to that though, in subsection 2.1 we lay down selected theories and empirical evidence on collaboration in the business context, which constitute a set of six theoretical arguments validating the conception of human-AI collaboration.

We argue that understanding the concept of collaboration is fundamental for conceptualization of human-AI collaboration and subsequently for generative artificial experts. Though collaboration by definition happens between people (American Psychological Association, 2024), we make a case that one can apply the knowledge about collaboration between people to collaboration between people and artificial intelligence agents (within human-AI collaboration).

3.1 Collaboration in The Business Context - Theories

Collaboration is a crucial aspect of the business landscape, as it stimulates sharing of knowledge, resources, and skills among individuals and organizations. By engaging in collaborative endeavors, businesses can unlock new opportunities for growth and innovation (Jassawalla & Sashittal, 2006). The following is a set of six theoretical arguments supporting emergence of human-AI collaboration and consequently generative artificial experts as collaborative systems, and how these theories inform development of the concept.

3.1.1 Collaboration Is Natural for People

Collaboration, being a mechanism resulting from evolutionary forces, happens naturally for people (Hill, 2002). Peoples' interdependence within groups led to development of cultural conventions and norms based on collaboration, setting our species aside as intrinsically collaborative (Tomasello et al., 2012). Thus, it is possible that human-AI collaboration may arise from these natural predispositions (Jemielniak & Przegalinska, 2020). Consequently, the benefits of collaboration between people and AI agents, as opposed to competition, may hold, as they do when people collaborate (Sowa et al., 2021). Since the understanding of human-human collaboration is extensive, it ought to be fundamental for considerations about human-AI collaboration. Furthermore, the natural inclination of humans towards collaboration is fundamental to GAEs, which are intended to work alongside humans in a collaborative manner.

3.1.2 People Perform a Cost-Benefit Analysis of Collaboration Under Social Exchange Theory

Social Exchange Theory (American Psychological Association, 2018) how humans engage in cost-benefit analyses when entering an interaction, and it is highly applicable to the case of GAEs. Users are likely to engage more with GAEs when the perceived benefits, such as enhanced productivity and decision-making, outweigh the costs. This theory underscores the importance of designing GAEs to offer tangible value in collaborative efforts.

3.1.3 Collaboration Outweighs Competition, Under Social Interdependence Theory

Social Interdependence Theory is a theoretical framework that directly supports the idea that collaborative (cooperative) efforts in certain contexts can lead to more beneficial outcomes than competitive strategies (Deutsch, 1949; Johnson & Johnson, 1989). This theory suggests that the way in which goals are structured within groups can significantly affect the outcomes of those groups, with cooperative (collaborative) goal structures often leading to more positive outcomes compared to competitive or individualistic structures. Positive interdependence exists when individuals working in tandem reach their goals together. In competition, people waste resources trying to outdo each other, while in collaboration resources are shared for a common goal, making this approach much more productive in the long-term. Furthermore, collaboration can provide emotional satisfaction and a sense of belonging. When individuals or groups collaborate, they can experience a sense of shared purpose and achievement, which can be emotionally rewarding (Nowak & Highfield, 2011). The value of collaboration over competition is mirrored in the design of GAEs. These systems are developed to collaborate with humans, sharing resources and expertise for a common goal, thereby enhancing overall productivity and fostering innovation, much more effectively than competitive approaches.

3.1.4 CPM Model Informs the Design of GAEs

Collaboration Process Model (CPM) (Mattessich & Monsey, 1992) offers a structured framework to analyze and comprehend the complexities of collaboration. According to the model, there are nineteen factors that influence the success of collaboration. The model assumes managing all the factors simultaneously, however in the context of this paper we highlight three: (1) *mutual respect, understanding, and trust*; (2) *clear roles, policy guidelines;* (3) *multiple layers of decision-making*. Collaboration thrives when participants recognize and value each other's expertise and more importantly - trust each other, fostering an environment where open communication and reliance. Well-defined responsibilities and expectations for all collaborators ensure that everyone understands their role and the rules governing the collaboration. The presence of different levels of decision-making authority within the collaborative process allows for a more nuanced and adaptive approach to problem-solving and decision-making, further enhancing the effectiveness of the collaboration. The factors of mutual respect, clear roles, and multiple layers of decision-making in CPM need to be reflected in the design of GAEs, which must respect human expertise, have well-defined functions, and be capable of making decisions (within boundaries) at different levels of autonomy within collaborative settings.

3.1.5 Collaboration Enhances Innovation and Knowledge Transfer

Collaboration in business may take various forms, ranging from informal knowledge sharing to more structured partnerships and alliances. For instance, cross-functional teams can benefit from the exchange of expertise and resources (Jassawalla & Sashittal, 2006). Inter-organizational collaboration is also positively linked to companies' performance (Cao & Zhang, 2011). Other benefits include increased innovation, improved product quality, cost saving in some cases. The same factors that influence the effectiveness of knowledge transfer between organizations, are also relevant when considering collaboration among individuals, as concepts of collaboration and knowledge transfer apply there as well. By engaging in collaborative efforts, individuals can expand their knowledge, enhance their skills, and increase their problem-solving capabilities, ultimately benefiting their personal and professional growth (Nonaka & Takeuchi, 1995). By integrating into various domains and working alongside human experts, GAEs facilitate the sharing of knowledge and foster innovative solutions, significantly benefiting from the diverse perspectives and expertise found in collaborative environments.

3.1.6 Collaboration Improves Productivity

Collaboration is essential for entrepreneurs, particularly in fostering innovation and leveraging diverse skills to increase efficiency. By engaging in collaborative efforts, entrepreneurs can tap into the collective intelligence of others, enabling them to generate new ideas and create innovative solutions. Entrepreneurs with more extensive and diverse networks are better able to generate high-quality innovation ideas than those with more limited or homogeneous networks. By actively seeking out and leveraging diverse network connections and collaboration, entrepreneurs can enhance their innovation capabilities (Björk & Magnusson, 2009). In industry 4.0, collaboration can act as an enabler and driver for increased productivity, provided that companies apply certain collaborative practices (Schuh et al., 2014). When technology is added to the mix, studies found interdependence between collaboration, technology and productivity (Sanders, 2007). Similarly, on the level of individual employees or entrepreneurs, technologies (social networks in this case) were found to increase collaboration between professionals and therefore lead to their higher productivity (Ferreira & Du, 2009). Though the consensus is that collaboration has a positive effect on productivity, a collaborative overload may bring the opposite effect (Lansmann & Klein, 2018).

3.2 Human-AI Collaboration (HAIC)

While human-AI collaboration is still an emerging concept, there has been a strong uptake in scientific publications featuring it (Hemmer et al., 2023). We start with a highlight of human-computer interaction, which we consider as step one, then human-AI interaction - step two, and then onto step three - human-AI collaboration. The conceptualization of generative artificial experts and their application in knowledge work relies fundamentally on human-AI collaboration.

3.2.1 From Human-Computer Interaction To Human-Ai Collaboration

Human-computer interaction (HCI) is a well-grounded field of research focused on interaction between people and technology, traditionally addressed by design and evaluation of computer interfaces. The emphasis lies in usability, ergonomics, and user experience. The field has evolved significantly since its inception, broadening to include interdisciplinary studies that encompass psychology, design, and social sciences (Kim, 2015). This process has led to increasingly sophisticated interfaces, both in hardware and software, aimed at enhancing the user's interaction with computer systems. These include interfaces employing natural language processing, such as chatbots or voice assistants, where rather than clicking buttons or sliders on a graphical user interface people use natural language to communicate their needs to the machine.

Advances in AI and subsequently increasing quality of natural language interaction forced emergence of human-AI interaction as a new construct. Human-AI interaction follows all key principles of HCI, as well as several new guidelines, applicable only to specificity of interaction with systems based on artificial intelligence (Amershi et al., 2019).

Using language is natural for humans, and so is collaboration (Hill, 2002). In the high-quality, natural language-based interaction between people and AI, collaboration between them is plausible, and would be a natural, next step. It goes beyond the simplified back and forth queries or commands characteristic for insofar human-AI interactions, entering a new space where both parties (human and AI agents) adapt, learn from interacting with each other. Simple interaction denotes a one-sided or transactional engagement - one party (human) needs something (e.g. an information or a solved task) and queries the other party (AI) for it. On the other hand, collaboration implies a more mutualistic relationship - that is both parties working together towards the same goal.

What may also draw people into collaborative endeavors with AI is their usual tendency to seek self-benefit from any interactions (Social Exchange Theory). As such, collaboration with AI has potential to bring several benefits to people, some of which have already been proven in experimental studies (e.g. increased productivity or possibility to automate selected tasks), and many are yet to be researched (Campero et al., 2022).

Collaborative intelligence is another theoretical construct that further substantiates the rationality of human-AI collaboration. While traditionally it focused on networks of people working together to solve a common problem with technology acting as an enabler, it has already been established that AI can become an intrinsic part of the “multi-agent problem-solving networks”, working hand-in-hand with people (Gill, 2012). Daugherty and Wilson (Daugherty & Wilson, 2018) extensively address the idea in their book, arguing that it is more likely for people to work together with artificial intelligence based systems rather than to be replaced by them in workplaces. Another substantial contribution to the field was Thomas Malone’s “Superminds” (Malone, 2018), showcasing the possibility for humans and computers to think and solve problems together. He distinguished four roles that machines can take towards people in such collaboration – ‘tools’, ‘assistants’, ‘peers’ and ‘managers’. All of them provide value to the user, as well boosts the teams’ performance, but on various levels of proximity. With 'tools' the interaction is one-sided, always initiated by a human. 'Assistants' act more autonomously, though within limits, and alike 'tools', their actions are fully controlled by human users. 'Peers' have a greater level of autonomy, yet always working at least under human supervision, they seamlessly interact with their user, as well as other parties. 'Managers' will possibly distribute and manage workload in mixed human and AI teams. Malone also argues that people and machines working together are examples of collective intelligence. AI can handle tasks such as data collection, pattern recognition, analyses, while humans can manage tasks that require emotional intelligence, empathy, abstract thinking, or ethical considerations. This division of labor can maximize the strengths of both entities and lead to a greater result than if they were working separately. As Malone argues, the role of such collaborations in driving innovation is emphasized to achieve outcomes neither could achieve alone. The concept of collaborative intelligence, where AI becomes a part of problem-solving networks alongside humans, is central to the idea of GAEs. GAEs are designed to act as assistants, peers, or even managers within these networks, complementing human intelligence with their data processing and analytical abilities, and contributing to a collective intelligence that surpasses individual capabilities.

Wang et al. (2020) call for a shift in focus from mere interaction to true collaboration between humans and AI, proposing a framework for Computer-Supported Cooperative Work to enhance mutual goal understanding and task co-management. Dellerman and colleagues (2021) identify the need for socio-technological ensembles that combine human and artificial intelligence, offering a structured design taxonomy to guide the development of systems that can achieve complex goals through hybrid intelligence. Nah and Zheng (2023) highlight the transformative potential of generative AI across diverse sectors including business, healthcare, and education, emphasizing the necessity for human-centered approaches to address ethical, technical, and regulatory challenges. Lastly, Memmert (2023) explores the dynamics of human-AI collaboration in brainstorming tasks, revealing that while generative large language models like GPT-3 can stimulate cognitive processes, they also include risks such as free riding, underscoring the complexity of integrating AI into traditional human group settings. This body of literature underscores the evolving landscape of human-AI interaction and sets a foundation for the introduction of generative artificial experts.

3.2.2 Examples of Human-AI Collaboration

Just as human-human collaboration in knowledge work can take many forms, so can human-AI collaboration, as demonstrated by several examples. The growing corpus of recent literature on the topic provides several case studies showcasing the versatile application areas of human-AI collaboration.

In an experimental study, human-AI collaboration in the field of marketing has been proven to bring better results in terms of productivity and work satisfaction, as compared to people working on their own (Sowa & Przegalinska, 2020). Other examples in knowledge work include possible usage of HAIC in disseminating expert knowledge in organizations and training new employees (Spitzer et al., 2022). Wang et al. (2023) presented an instance of employing HAIC for data science work. They demonstrated a process where a human worker and AI work together on solving a problem, handing over the task between them several times. First, the human prepares grounds for a data analysis, selects key data points and appropriate visualization layouts, then the AI system performs an analysis and creates slides. Later, the human judges the outcome, proposed refinements, and customizations, which are handled by the AI system.

A strong case has been made for HAIC in healthcare, where it could help address a problem of qualified staff shortages, increase specialists' productivity, and benefit the overall patient care. Though there are many challenges to address, such as transparency of decision-making, bias, data security, further research is needed (Lai et al., 2021). Optimal model for human-AI collaboration in medical decision-making is still in search, though without a doubt the emergence of the phenomena is beneficial to patients. In an experimental study of analyzing colonoscopy videos, “human-AI hybrid teams” outperformed humans alone or AI agents tackling the same task on their own (Reverberi et al., 2022).

In a physical world, truck drivers were proven to be more efficient in route planning when they were teamed up with an assistive AI. In this case, human-AI teams were found to achieve better results than humans only, even though they were experienced in the field (Loske & Klumpp, 2021). Another tangible example are cobots - collaborative robots, which solve complex tasks in factories “hand-in-hand” with people (Colgate et al., 2023). Saffiotti et al. (2020) proposed a model of human-AI collaboration in arts and music performances - combining human pianists with AI drummers, human dancers with AI drummers and human pianists with robot dancers.

These examples present the potential of human-AI collaboration across diverse fields, demonstrating its ability to enhance productivity, improve decision-making, and tackle complex tasks more effectively than humans or AI alone. The breadth of applications highlights the importance of continued research on the topic but also warrants a wide spectrum of possible applications for GAEs.

4. Generative Artificial Experts in Detail

This section addresses the definition of generative artificial experts and provides practical insights into the concept. In the first part we provide a definition of GAEs and list seven defining traits of GAEs. In the second part we present illustrative examples of GAEs which help to understand their applicability and later discuss how existing technologies relate to GAEs.

4.1 Generative Artificial Experts - Definition

A generative artificial expert is a collaborative generative AI agent designed for human-AI collaboration within a specialized domain. GAEs are based on multimodal generative AI, perform actions, work within bounded autonomy, among other features, to actively participate in problem-solving and decision-making processes with people. GAEs harness the capabilities of generative AI in a manner that prioritizes collaboration and enhances human work. These agents are designed to interact with users using conversational interfaces, providing targeted assistance in solving complex and specialized tasks that require expert-level knowledge. This approach distinguishes GAEs from other AI systems where generative components are just one part of a larger, less focused system.

What differentiates GAEs from generic LLM-based systems such as ChatGPT, Gemini or Claude is their in-depth understanding of a selected, specialized domain. This specialization is achieved through training on domain-specific data and scenarios. Another differentiator is the ability to operate semi-independently within user's workflow, make certain decisions, and execute a range of tasks.

4.2 Defining Traits of GAEs

The defining traits of GAEs is a list of 7+1 elements which provides a detailed understanding of the concept and sets them apart from other AI-based systems. The combination of all these seven traits contribute to the definition of the concept. The defining traits were described in Table 1. A system to be considered a GAE needs to fulfill all the listed traits. It is possible that some systems, even existing ones, might resemble GAEs - achieving some of the traits, but not all of them. We would refer to these kinds of systems as *quasi-GAEs,* if they were applied in a GAE context.

Table 1. Defining traits of generative artificial experts

Defining trait	Description	Implementation perspective
Collaborative Attitude	Acting as collaborative partners that enhance productivity by working alongside humans.	Ensure that communication style and approach to problem-solving is collaborative. GAEs anticipate needs, proactively provide support, and discuss solutions.
Domain Expertise	Possess specialized knowledge in a specific field, applying this expertise effectively to specific problems with precision, in a way suitable for a given domain.	A common knowledge base gives the system ability to reason on a general level. Specialist expertise is achieved through training of domain-specific data.
Synthetic Persona	Could be designed with crafted personas (e.g. ‘a mature professor’ or ‘laid back entrepreneur), aligning their communication style and interaction to resonate with the cultural and professional context of users.	Incorporating language models with adjusted tone, formality, and terminology based on user preferences and the professional context.
Multimodal Capabilities	Process and generate various data forms, adapting to user needs and task requirements by switching or combining modalities for a richer experience.	Enable multimodal capabilities by integrating models that allow GAEs to process and generate textual data, images, video, code, audio and other modalities.
Personal or Organizational Fit	Seamlessly embedded in workflows. Can be customized to align with specific user or organizational needs, e.g. in terms of terminology, work patterns, organizational knowledge.	Achieve personal or organizational fit by enabling users to teach GAEs, e.g. about user-specific workflows or terminology.
Bounded Autonomy	Operate within predefined autonomy, executing selected tasks independently within set parameters while maintaining alignment with human oversight and objectives.	Define bounded autonomy by setting up triggers and constraints within the GAE system, allowing it to perform some tasks independently while flagging exceptions for human review.
Performing Actions	Perform a range of actions within the user's workflow, interacting with various systems to proactively complete tasks.	Facilitate action performance by connecting GAEs to various enterprise systems and databases, enabling them to execute tasks such as data entry, file management, and system monitoring autonomously.
Ethical consideration	Require users to ensure operations align with ethical standards, maintaining oversight and accountability for all generated content and decisions.	Implement ethical safeguards by embedding compliance checks and ethical guidelines within the GAE’s decision-making algorithms, ensuring all actions align with ethical, organizational and legal standards.

4.3 Illustrative Examples of GAEs

We employ the method of illustrative examples to showcase generative artificial experts in practice (Rawson et al., 2015). It involves the creation of detailed, hypothetical examples that are designed to simulate the practical application and outcomes of GAEs, due to the absence of existing real-world examples. They are described in detail in this section, and positive and negative aspects related to GAEs were addressed later in the paper.

4.3.1 Finley - Financial Trading Analyst

Finley, a generative artificial expert, is developed for applications in the financial sector, particularly concentrating on stock market data and investment strategies. This GAE supports employees of corporate financial institutions, such as trading teams in consulting companies, banks or funds. The primary function of Finley is to analyze and interpret market trends and stock exchange data, guiding investment decisions and trade executions. The base knowledge includes understanding of financial theories and practical methodologies in asset management and investment analysis. Its knowledge and skills were confirmed by the passing financial certification tests such as CFA, CFP, CPA and CAIA. Finley is capable of processing real-time market data, generating and verifying investment strategies, and analyzing present and past market trends.

Table 2. Defining traits of Finley GAE

Defining trait	Description
Collaborative Attitude	Engages in active discussions on investment strategies, their potential gains and costs. Is flexible to act as an advisor or to execute trade requests from a user, depending on which behavior is necessary.
Domain Expertise	Comprehensive database of market data, understanding of financial instruments and concepts; real time update with market research and news. Knowledge confirmed with passing of CFA, CFP, CPA and CAIA certifications.
Synthetic Persona	Professional, data-driven communication style; adjusts complexity based on user's expertise, varying from technical to educational. Adopts a tone that's assertive and jargon-heavy, reminiscent of a Wall Street veteran, often using complex financial terms and industry slang that resonates with the 'investment club' ethos, conveying market savvy and insider knowledge.
Multimodal Capabilities	Generates and interprets financial charts, graphs, and heat maps; converts text reports into spoken briefings for auditory updates. Able to run data analysis and interpret results.
Personal or Organizational Fit	Configurable to match user-specific investment strategies, risk profiles, and regulatory environments; adapts to different user needs from investment funds to individual traders.
Bounded Autonomy	Autonomously monitors market data, alerts users, and executes trades within set parameters; maintains alignment with user's risk tolerance and strategy.
Performing Actions	Executes financial tasks beyond analysis, including automated trading, portfolio rebalancing, and interfacing with financial databases for data gathering.

4.3.2 FlowMaster - DevOps Coordinator GAE

FlowMaster, as a DevOps Coordinator GAE, is designed to optimize the DevOps processes in software development environments. Its primary function is to automate and refine various stages of the software development lifecycle, including code integration, testing, deployment, and subsequent monitoring phases. In its alignment with CI/CD methodologies, FlowMaster integrates with pre-existing infrastructures, through automated code merging, testing, and deployment processes. The GAE evaluates existing workflows, identifying and addressing inefficiencies, thereby optimizing both speed and resource allocation. Post-deployment, it engages in performance monitoring, providing feedback for the refinement of development processes.

Table 3. Defining traits of FlowMaster GAE

Defining trait	Description
Collaborative Attitude	Proactively suggest solutions to problems it discovers.
Domain Expertise	In-depth understanding of DevOps lifecycle; knowledgeable in coding, testing, deployment strategies, and system monitoring.
Synthetic Persona	Communicates effectively with developers and operations staff. It is proactive in communication. Adopts the persona of a DevOps expert, communicating with technical precision and clarity, using a language that resonates with both developers and operations staff, facilitating effective and efficient project collaboration. Tone is methodical and authoritative, often using structured, concise language that conveys a sense of expertise and reliability.
Multimodal Capabilities	Processes and generates data in various forms including text, code, charts, and audio alerts, ensuring effective communication in preferred formats.
Personal or Organizational Fit	Seamlessly integrates with DevOps tools; automates code merges, testing, and deployments. Adapts to specific organizational DevOps practices; aligns with coding standards, deployment practices, and operational policies for personalized integration.
Bounded Autonomy	Executes tasks with bounded autonomy like initiating deployments and managing resources, under human supervision to align with organizational goals and compliance.
Performing Actions	Actively performs tasks in DevOps cycle; manages deployments, cloud resources, automated testing, and applies updates, directly contributing to software development.

4.3.3 Muse - Creative Director GAE

Muse, the GAE, has the ability to emulate the role of a creative director in a small business organization, with a particular focus on advertising and product design. Central to Muse's capabilities is the ideation, generation and review of campaign concepts, brand identity, product designs, and marketing content. Muse is proficient in crafting both visual elements and copy, tailored to align with the thematic essence of campaigns and the overarching brand voice. This capability extends to the production of diverse multimedia content, including digital graphics, video concepts, and interactive web designs, positioning Muse as a comprehensive creative tool. Muse can work alongside creative or marketing teams, or act as a consultant for companies which do not have these teams yet.

Table 4. Defining traits of Muse GAE

Defining trait	Description
Collaborative Attitude	Acts as a consultant for firms without in-house creative teams or a brainstorming collaborator to those that do.
Domain Expertise	Expertise includes an in-depth understanding of visual aesthetics, knowledge of art, design and branding principles, storytelling, campaign conceptualization, and audience engagement strategies. Has the ability to create and maintain a creative vision
Synthetic Persona	Embodies the persona of an experienced and insightful creative director, articulate in conveying complex creative concepts with clarity and engaging in thoughtful dialogue, often using vivid, imaginative language to inspire and guide teams. Uses a tone that is vibrant and imaginative, often weaving creative expressions and storytelling elements into its communication.

Defining trait	Description
Multimodal Capabilities	Generates a range of creative outputs, from text and visuals to multimedia, adaptable to various project needs.
Personal or Organizational Fit	Integrates with existing creative processes and teams, contributing to ideation, strategy formulation, and feedback. Adapts to the unique creative, stylistic, and branding requirements of different organizations, ensuring alignment with each brand's identity and campaign objectives.
Bounded Autonomy	Operates within set strategic and creative boundaries, autonomously generating concepts and content while adhering to campaign goals and brand values.
Performing Actions	Directly contributes to the creation of advertising materials, drafting copy, designing visuals, and producing multimedia content.

4.4 Existing Technologies and GAEs

As of now (Q1 2025), there are no existing systems that meet all defining traits of GAEs. However, with the rapid development of generative AI and AI agents, it is likely that soon such technologies will exist. In this section we briefly describe existing systems that that include some defining traits (quasi-GAEs), which could be a starting point to develop fully functioning generative artificial experts.

"GPTs" created by OpenAI and launched in November 2023 within their product ChatGPT are a close representation of what GAEs could be. GPTs are "custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills" (OpenAI, 2023). Some examples of GPTs include "Academic Assistant Pro", which supports users in writing papers, data interpretation and other research related tasks, "DesignerGPT" which generates website designs and writes HTML code for them. GPTs have some domain expertise, although they are not trained on domain-specific data, they have multimodal capabilities and could have a persona. However, their ability to integrate in collaborative workflows, personalized fit, performing actions and bounded autonomy are limited or not possible currently.

Another link from existing technology to GAEs can be found in multi-agent systems. These systems have received increasing attention in recent years, especially as generative AI becomes more capable and more widely adopted. Multi-agent systems are composed of several AI agents aligned together, where each is designed to perform a specific task and is equipped with the necessary capabilities to execute them effectively (Dorri et al., 2018). For example, Li et al (2023) successfully developed a multi-agent system for financial trading. Using several LLM-based agents, authors aimed to better emulate human cognitive processes, compared to a single LLM-based agent, therefore making the system more useful in decision support for human traders.

Deep research agents, such as those developed by OpenAI, Perplexity, Google, and X, are another example. These systems are built to browse the web intelligently, searching through text, images, and documents to find, analyze, and summarize useful information. They use advanced models designed for web research and data processing, allowing them to quickly adjust as new information comes in. They can create detailed, well-sourced reports in any domain, similar to what a professional researcher would produce (OpenAI, 2025). In terms of GAEs similarity and differences, they execute tasks with bounded autonomy and provide users with reports that include a synthesis of findings. They are somewhat collaborative, i.e. they typically ask users for directions before executing the research task. However, despite being able to produce a domain-specific report, they themselves do not have the domain-specific training. Their scope is currently limited to online search, and they do not have a synthetic persona.

5. Impact and Discussion

This section discusses the positive and negative impacts that generative artificial experts may bring. Since GAEs do not exist in the real world yet, any claim related to its impact would be speculative. Instead, we use analogical reasoning to draw connections from fields of research of technologies that share fundamental similarities with GAEs, allowing us to apply insights from a well-understood area to make possible predictions. The true impacts of GAEs remain unknown until such systems are developed and empirically evaluated.

5.1 Positive Impacts of GAEs

5.1.1 Productivity increase

Generative AI has already demonstrated significant productivity gains across various organizational settings (Dell'Acqua et al., 2023; Noy & Zhang, 2023; Reverberi et al., 2022; Sowa & Przegalinska, 2020). Generative artificial experts will likely extend this capability, by providing targeted, domain-specific support to users. Users may decide to delegate mundane or routine tasks, or offload some cognitive processes, creative tasks or engage in brainstorming (Epstein et al., 2023; Platt & Platt, 2023). In any case, their productivity is likely going to increase (Brynjolfsson et al., 2023).

5.1.2 Democratization of knowledge

Generative AI systems, such as ChatGPT. facilitate the democratization of knowledge by making it easily available to non-experts (Bilgram & Laarmann, 2023; Brynjolfsson et al., 2023). GAEs could do the same, however with disseminating expert-level knowledge (harder to come by) to individual workers. This accessibility allows employees to improve their work performance and provides them with self-development opportunities. For example, a GAE could provide real-time guidance to a junior researcher conducting advanced scientific experiments, enhancing their capability to operate at a higher level and accelerating their learning curve. It does not mean that a junior researcher teamed up with a GAE could outperform a human research expert, however it would mean that the outcome of the work could be better, and the junior researcher would make less mistakes.

5.1.3 Accessing capabilities that require significant resources

Adopting any innovative and powerful technology might give small organizations a competitive edge (Makkonen, 2008). Generative artificial experts could significantly empower smaller organizations by providing them with expert-level capabilities that are usually prohibitive due to costs or the scarcity of specialized human resources. An example could be a small company using a GAE for financial forecasting, tasks that would traditionally require the hiring a consultant.

5.1.4 Enhanced decision-making

Studies on decision support systems have shown their effectiveness in enhancing decision-making accuracy across various domains (Sharda et al., 1988). Generative artificial experts build on this foundation, potentially offering even greater precision and adaptability in decision-making processes. An example could be a GAE used in healthcare to assess patient data and recommend personalized treatment plans, thereby increasing the accuracy and personalization of medical care.

5.1.5 Reducing model hallucinations

Generic generative AI systems often produce hallucinations, particularly when dealing with niche domains, leading to unreliable or inaccurate outputs (Li et al., 2023). Generative artificial experts are specifically designed to address this issue, promising a significant reduction in hallucinations by leveraging training on specialized, high-quality datasets tailored to specific domains.

5.2 Negative Impacts, Concerns and Challenges of GAEs

5.2.1 Possible job loss

The introduction of generative artificial experts in the workplace raises concerns about job displacement. However, while GAEs can automate certain tasks, their primary role is to support human capabilities, not replace them. Bessen (2019) suggests that work automation may necessitate disruptive transitions for workers, who must adapt to new industries requiring new skills. Tschang and Almirall (2021) elaborate on the dual nature of AI effects on the job market, noting that while AI can augment jobs and reduce the demand for routine skills, it may also transform the balance of work, potentially reducing the quantity of high-skilled roles over time. The expected impact of GAEs on the labor market is no different than that expected and already studied in relation to a wide adoption of AI technologies, particularly generative AI. That being noted, GAEs are expected to be niche systems, rather than mainstream, thus their macroeconomic impact might be limited.

5.2.2 Dependence on technology

In a study about marine navigation Wu and colleagues (2022) clearly express that "over-reliance on technologies in (...) may have disastrous consequences". An over-reliance on GAEs could lead organizations to become overly dependent on technological solutions, potentially at the expense of human expertise and judgment, or making ruinous mistakes. This dependence could make businesses vulnerable to system failures or cybersecurity attacks, where significant portions of their operations could be disrupted due to issues with their AI systems. For instance, an investment firm that relies heavily on a GAE for financial forecasting and analysis, might be led to financial loss in case the system's predictions prove wrong.

5.2.3 Creation of echo chambers

Echo-chambers are studied in the context of social media platforms, where users are exposed primarily to information and opinions that reinforce their existing beliefs, often limiting exposure to diverse perspectives and creating polarized communities (Terren & Borge-Bravo, 2021). GAEs might become such chambers and reinforce existing ideas without offering critical or diverse perspectives. A GAE that would be designed to be agreeable and supporting will simply repeat and agree to notions proposed by human users. For example, A GAE used for legal research might prioritize delivering case law that aligns with a lawyer's previous arguments, potentially overlooking important contrary precedents.

5.2.4 Skill degradation

A study of flight planners has shown a cognitive skills degradation of workers, as a result of task automation (Volz & Dorneich, 2020). As GAEs take over more complex and varied tasks, there's a risk that human skills could degrade over time. Workers may rely too heavily on AI for decision-making, problem-solving, and creative processes, potentially leading to a decline in critical thinking and problem-solving abilities in the workforce. For instance, engineers who rely on a GAE for troubleshooting machine performance, might gradually lose their own ability to diagnose and solve complex problems without AI assistance.

5.2.5 Reduction in Human Agency

Increased use of algorithmic management in gig economy platforms has been critiqued for diminishing worker autonomy by dictating work patterns and behaviors. While this led to some desirable outcomes (e.g. increased flexibility, autonomy of workers), it also led to lower pay, social isolation and overwork, among others (Wood et al., 2019). There is a concern that as decision-making processes become more automated and reliant on GAEs, human agency could be diminished. This could lead to a scenario where individuals feel less in control of their decisions, potentially impacting mental health and job satisfaction. To provide an example, we can imagine a consulting firm using a GAE to assign tasks and evaluate employee performance, leading employees to feel they have little control over their work choices or career progression.

5.2.6 Exploitation by bad actors

The use of generative AI for creating deep-fakes has been documented in creating false media content, impacting public opinion and privacy (Chesney & Citron, 2019). The advanced capabilities of generative artificial experts could be exploited by malicious entities. Bad actors could use these systems to enhance their efforts in cyber-attacks, fraud, and misinformation campaigns. For example, GAEs could be used to create sophisticated phishing emails that are indistinguishable from legitimate communications or to generate fake news articles and deep-fake videos that could sway public opinion or manipulate financial markets. The potential for such misuse raises significant security concerns and necessitates robust mechanisms to prevent unauthorized access and misuse of these powerful tools.

5.2.7 Evaluation of expertise

Evaluating and validating the expertise of generative artificial experts is a crucial challenge to address. In traditional human settings, an 'expert' is often validated through a mixture of formal credentials, demonstrated experience, peer recognition, and contributions to the field. These methods, however, are not applicable to generative artificial experts, necessitating the establishment of alternative validation mechanisms to ensure their reliability and relevance. We call for a strict control over the training data used in developing these systems, with a strong emphasis on transparency throughout the training process, per the nature of the data and specific application areas. Reinforcement learning with feedback should prove particularly useful as a method for refining the outputs of these AI systems based on human expert evaluations. To robustly address the validation of domain knowledge and expertise embedded in these systems, we recommend three primary approaches: human-grade assessments such as examinations or certifications (wherever feasible), real-world scenario testing supplemented by human expert feedback, and known technical evaluations of AI model performance (e.g. benchmarking, knowledge probing, domain-specific assessments, reasoning tests, bias evaluations, and robustness testing). Explainability of AI systems is a known research problem that also concerns GAEs. It is vital that the results of any evaluation of a generative artificial expert's expertise be transparent to its users, as it fosters trust in the system especially among users with high domain expertise (Bayer et al., 2022). Should these evaluation methods prove insufficient, this area would undoubtedly require further in-depth research to develop more effective validation mechanisms.

5.2.8 Summary of challenges

The deployment of GAEs necessitates a strong focus on ethical considerations. This includes ensuring data privacy, avoiding biases in AI decision-making, and maintaining transparency in AI processes, and above all - ethical application of GAEs. Responsible use of GAEs involves regular monitoring for ethical compliance and the establishment of guidelines to govern their application. For example, the IEEE’s Ethically Aligned Design could be used as an ethical framework to guide discussion and development of GAEs. This framework, particularly its principles related to human rights and data agency, is critical for shaping responsible AI practices and can be adapted to guide the creation and application of GAEs. Building on this, a tailored set of ethical guidelines should be developed, emphasizing transparency, fairness, and accountability throughout the lifecycle of GAEs. Additionally, the necessary measures should be taken to ensure GAEs transparency within explainable AI frameworks, to make GAE decisions understandable and justifiable for users. Practical measures to enhance transparency include incorporating user-friendly interfaces that explain AI decisions in simple terms and ensuring training documentation and system updates for all stakeholders, thus fostering a deeper trust and broader acceptance of GAE technology.

5.3 Discussion

In this paper we devised a path for emergence of generative artificial experts, and proposed a definition of this concept, together with defining traits which set them apart from other generative AI systems. We highlighted possible positive and negative impacts, using analogical reasoning based on other technologies. The discussion section primarily focuses on issues that we observed along the way that are related to GAEs and that remain open and require further exploration and understanding.

Based on our findings from analyzing existing research on human-AI collaboration, there is a need for a revised theoretical framework of collaboration in general, one that specifically addresses human-AI collaboration. Current collaboration models primarily focus on human-to-human interactions and may not fully capture the unique challenges that AI introduces, such as cognitive dissonance related to accepting AI-driven insights, communication barriers, ambiguity in role hierarchy, power dynamics etc. Therefore, developing a theory that collaboratively integrates humans and AI agents is crucial for understanding and optimizing these new forms of interaction.

Addressing the possible negative impacts of GAEs mentioned in Section 5 is paramount. So are ethical and societal implications related to GAEs. Ensuring ethical development and deployment of GAEs will require collaborative efforts among technologists, ethicists, policymakers, and other stakeholders.

The technological feasibility and challenges of creating GAEs also warrant attention. The technical side of development of these advanced AI systems has not been addressed yet. Further technical research and development is required to prove feasibility of developing GAEs in the described form.

Another issue related to the ‘expertise’ of generative artificial experts is the validation of their knowledge and expertise. For now, authors propose using a combination of methods such as human-grade assessments (e.g. certifications, evaluations by independent judge panels, or exams), real-world scenario testing with feedback from human experts, and technical evaluations of the AI model's performance. If these methods prove insufficient, there is a significant opportunity to develop new standards and techniques for evaluating AI-driven knowledge and expertise. In any case, generative artificial experts should always undergo transparent validation before their real-world application to ensure reliability and accuracy.

While this study establishes a foundational understanding of Generative Artificial Experts, it does not discuss strategies for managing organizational change or user adoption of GAEs. These are significant topics that warrant detailed exploration in future studies to fully understand their implications in practical settings.

In conclusion, while GAEs present a future brimming with possibilities, realizing this potential necessitates continued research. As this concept evolves, it is vital to proactively address the challenges and opportunities that GAEs bring to the forefront.

6. Conclusion

This paper provided a conceptualization of generative artificial experts. It provided a comprehensive definition of the proposed system, including the defining traits that sets them apart from typical generative AI systems. Several possible avenues for development were discussed and illustrative examples were employed to present GAEs' potential applications. Possible positive and negative impacts of GAEs were also highlighted, including enhanced decision-making and productivity, and echo chambers problem or impact on labor market. The article contributes to understanding GAEs' role as a possibly relevant system in the future of future knowledge work.

The future of GAEs, while promising, remains uncertain. The concept of GAEs, as outlined in this paper, offers a compelling vision of AI's potential but is still in its conceptual phase. Factors such as technological feasibility, system design, user acceptance, ethical standards alignment, will play significant roles in determining the timeline and extent of GAEs' development and adoption. These need to be addressed by future research.

Acknowledgements

We wish to express our gratitude to Aureliusz Gorski and CampusAI (https://www.campus.ai/) for introducing us to the idea of using the GPT-4 model to create generative artificial experts for startup companies. It is through his practical efforts in dedicating extensive hours towards this exploring this idea that we were able to develop the GAE concept for this research paper.

This research was funded by the Polish National Science Center (NCN) OPUS 21 grant no. 2021/41/B/HS4/03664.

References

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey. *IEEE Access*, *13*, 18912–18936. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3532853

Alavi, M., & Westerman, G. (2023, November 7). How Generative AI Will Transform Knowledge Work. *Harvard Business Review*. https://hbr.org/2023/11/how-generative-ai-will-transform-knowledge-work

American Psychological Association. (2018). *APA Dictionary of Psychology*. https://dictionary.apa.org/

American Psychological Association. (2024). *APA Dictionary of Psychology*. https://dictionary.apa.org/

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

Bayer, S., Gimpel, H., & Markgraf, M. (2022). The role of domain expertise in trusting and following explainable AI decision support systems. *Journal of Decision Systems*, *32*(1), 110–138. https://doi.org/10.1080/12460125.2021.1958505

Bessen, J. (2019). Automation and jobs: When technology boosts employment*. *Economic Policy*, *34*(100), 589–626. https://doi.org/10.1093/epolic/eiaa001

Bilgram, V., & Laarmann, F. (2023). Accelerating Innovation With Generative AI: AI-Augmented Digital Prototyping and Innovation Methods. *IEEE Engineering Management Review*, *51*(2), 18–25. IEEE Engineering Management Review.

https://doi.org/10.1109/EMR.2023.3272799

Björk, J., & Magnusson, M. (2009). Where Do Good Innovation Ideas Come From? Exploring the Influence of Network Connectivity on Innovation Idea Quality. *Journal of Product Innovation Management*, *26*(6), 662–670.

Borsos, Z., Marinier, R., Vincent, D., Kharitonov, E., Pietquin, O., Sharifi, M., Roblek, D., Teboul, O., Grangier, D., Tagliasacchi, M., & Zeghidour, N. (2023). AudioLM: A Language Modeling Approach to Audio Generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 2523–2533.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners*.

Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative AI at Work* (w31161; p. w31161). National Bureau of Economic Research. https://doi.org/10.3386/w31161

Campero, A., Vaccaro, M., Song, J., Wen, H., Almaatouq, A., & Malone, T. W. (2022). *A Test for Evaluating Performance in Human-Computer Systems*.

Cao, M., & Zhang, Q. (2011). Supply chain collaboration: Impact on collaborative advantage and firm performance. *Journal of Operations Management*, *29*(3), 163–180.

Chang, Z., Koulieris, G. A., & Shum, H. P. H. (2023). *On the design fundamentals of diffusion models: A survey*.

Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, *107*(6), 1753–1820.

Chu, H.-C., & Hwang, G.-J. (2008). A Delphi-based approach to developing expert systems with the cooperation of multiple experts. *Expert Systems with Applications*, *34*(4), 2826–2840.

Colgate, J. E., Wannasuphoprasit, W., & Peshkin, M. A. (2023). *Cobots: Robots for Collaboration With Human Operators*. 433–439. https://doi.org/10.1115/IMECE1996-0367

Crevier, D. (1993). *Ai: The Tumultuous History Of The Search For Artificial Intelligence*. Basic Books.

Daugherty, P. R., & Wilson, J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Review Press.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (SSRN Scholarly Paper 4573321). https://doi.org/10.2139/ssrn.4573321

Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2021). *The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems* (arXiv:2105.03354). arXiv. https://doi.org/10.48550/arXiv.2105.03354

Deutsch, M. (1949). A Theory of Co-operation and Competition. *Studies towards the Integration of the Social Sciences, 2,* 129–152.

Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-Agent Systems: A Survey. *IEEE Access*, *6*, 28573–28593. IEEE Access. https://doi.org/10.1109/ACCESS.2018.2831228

Epstein, Z., Hertzmann, A., & THE INVESTIGATORS OF HUMAN CREATIVITY. (2023). Art and the science of generative AI. *Science*, *380*(6650), 1110–1111. https://doi.org/10.1126/science.adh4451

Ferreira, A., & Du, P. T. (2009). Effect of online social networking on employee productivity. *South African Journal of Information Management*, *11*(1), 1–11. https://doi.org/10.10520/EJC46309

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Giarratano, J. C., & Riley, G. (2005). *Expert Systems: Principles and Programming*. Thomson Course Technology.

Gill, Z. (2012). User-driven collaborative intelligence: Social networks as crowdsourcing ecosystems. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (pp. 161–170).

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. The MIT Press.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). *Building Expert Systems*. Addison-Wesley Publishing Company.

Hemmer, P., Westphal, M., Schemmer, M., Vetter, S., Vössing, M., & Satzger, G. (2023). Human-AI Collaboration: The Effect of AI Delegation on Human Task Performance and Task Satisfaction. *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 453–463. https://doi.org/10.1145/3581641.3584052

Hill, K. (2002). Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature (Hawthorne, N.Y.)*, *13*(1), 105–128. https://doi.org/10.1007/s12110-002-1016-3

Hubert, & Dreyfus, S. (1997). Why computers may never think like people. In *Knowledge Management Tools* (pp. 31–50). Elsevier.

Jackson, P. (1999). *Introduction to Expert Systems*. Addison-Wesley.

Janik, R. A. (2023). *Aspects of human memory and Large Language Models*.

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, *61*(4), 577–586.

Jassawalla, A. R., & Sashittal, H. C. (2006). Collaboration in Cross-Functional Product Innovation Teams. In M. M. Beyerlein, S. T. Beyerlein, & F. A. Kennedy (Eds.), *Innovation through Collaboration* (Vol. 12, pp. 1–25). Emerald Group Publishing Limited. https://doi.org/10.1016/S1572-0977(06)12001-4

Jemielniak, D., & Przegalinska, A. (2020). *Collaborative society*. MIT Press.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and Competition: Theory and Research*. Interaction Book Company.

Kim, G. J. (2015). *Human-Computer Interaction: Fundamentals and Practice*. CRC Press.

Kumar, A., Lakshmi Devi, M. S., & Saltz, J. S. (2023). Bridging the Gap in AI-Driven Workflows: The Case for Domain-Specific Generative Bots. *2023 IEEE International Conference on Big Data (BigData)*, 2421–2430. https://doi.org/10.1109/BigData59044.2023.10386894

Lai, Y., Kankanhalli, A., & Ong, D. (2021). *Human-AI Collaboration in Healthcare: A Review and Research Agenda*. 1–10. https://doi.org/10.24251/HICSS.2021.046

Lansmann, S., & Klein, S. (2018). How much Collaboration? Balancing the Needs for Collaborative and Uninterrupted Work. *ECIS*, *118*.

Li, Y., Yu, Y., Li, H., Chen, Z., & Khashanah, K. (2023). *TradingGPT: Multi-Agent System with Layered Memory and Distinct Characters for Enhanced Financial Trading Performance* (arXiv:2309.03736). arXiv. https://doi.org/10.48550/arXiv.2309.03736

Lindsay, R. K., Buchanan, B. G., Feigenbaum, E. A., & Lederberg, J. (1993). DENDRAL: A case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence*, *61*(2), 209–261. https://doi.org/10.1016/0004-3702(93)90068-M

Loske, D., & Klumpp, M. (2021). Intelligent and efficient? An empirical analysis of human-AI collaboration for truck drivers in retail logistics. *The International Journal of Logistics Management*, *32*(4), 1356–1383.

Makkonen, H. (2008). A Small Business as Investor—A Single Case Study of a Technology Adoption. *Piccola Impresa / Small Business*, *1*, Article 1. https://doi.org/10.14596/pisb.104

Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together* (First edition). Little, Brown and Company.

Mattessich, P. W., & Monsey, B. R. (1992). *Collaboration–what Makes it Work: A Review of Research Literature on Factors Influencing Successful Collaboration*. Amherst H. Wilder Foundation.

Mccarthy, J. (1984). Some expert systems need common sense. *Annals of the New York Academy of Sciences*, *426*, 129–137.

Memmert, L., & Tavanapour, N. (2023, May 11). Towards Human-AI-Collaboration in Brainstorming: Empirical Insights into the Perception of Working with a Generative AI. *31st European Conference on Information Systems (ECIS)*.

Mukherjee, A., & Chang, H. (2023). *The Creative Frontier of Generative AI: Managing the Novelty-Usefulness Tradeoff* (arXiv:2306.03601). arXiv. https://doi.org/10.48550/arXiv.2306.03601

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

Nowak, M., & Highfield, R. (2011). *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. Simon and Schuster.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187–192. https://doi.org/10.1126/science.adh2586

OpenAI. (2023, November 6). *Introducing GPTs*. https://openai.com/index/introducing-gpts/

OpenAI. (2025, February 2). *Introducing deep research*. https://openai.com/index/introducing-deep-research/

Pal, D., Vanijja, V., Thapliyal, H., & Zhang, X. (2023). What affects the usage of artificial conversational agents? An agent personality and love theory perspective. *Computers in Human Behavior*, *145*, 107788. https://doi.org/10.1016/j.chb.2023.107788

Partridge, D. (1987). The scope and limitations of first generation expert systems. *Future Generations Computer Systems: FGCS*, *3*(1), 1–10.

Piller, F. T., Srour, M., & Marion, T. J. (2024). Generative AI, Innovation, and Trust. *The Journal of Applied Behavioral Science*, *60*(4), 613–622. https://doi.org/10.1177/00218863241285033

Platt, M., & Platt, D. (2023). Effectiveness of Generative Artificial Intelligence for Scientific Content Analysis. *2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT)*, 1–4. https://doi.org/10.1109/AICT59525.2023.10313167

Poole, D. L., & Mackworth, A. K. (2023). *Artificial Intelligence: Foundations of Computational Agents* (3rd ed). Cambridge University Press.

Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., & Cherubini, A. (2022).

Experimental evidence of effective human–AI collaboration in medical decision-making. *Scientific Reports*, *12*(1), 14952. https://doi.org/10.1038/s41598-022-18751-2

Rick, S. R., Giacomelli, G., Wen, H., Laubacher, R. J., Taubenslag, N., Heyman, J. L., Knicker, M. S., Jeddi, Y., Maier, H., Dwyer, S., Ragupathy, P., & Malone, T. W. (2023). *Supermind Ideator: Exploring generative AI to support creative problem-solving*.

Roychowdhury, S. (2023). *Journey of Hallucination-minimized Generative AI Solutions for Financial Decision Makers* (arXiv:2311.10961). arXiv. https://doi.org/10.48550/arXiv.2311.10961

Saffiotti, A., Fogel, P., Knudsen, P., De Miranda, L., & Thörn, O. (2020). On human-AI collaboration in artistic performance. In *First International Workshop on New Foundations for Human-Centered AI (NeHuAI) Co-Located with 24th European Conference on Artificial Intelligence (ECAI 2020)* (pp. 38–43).

Sanders, N. R. (2007). An empirical study of the impact of e-business technologies on organizational collaboration and performance. *Journal of Operations Management*, *25*(6), 1332–1347.

Schuh, G., Potente, T., Varandani, R., Hausberg, C., & Fränken, B. (2014). Collaboration Moves Productivity to the Next Level. *Procedia CIRP*, *17*, 3–8. https://doi.org/10.1016/j.procir.2014.02.037

Sharda, R., Barr, S. H., & McDonnell, J. C. (1988). Decision Support System Effectiveness: A Review and an Empirical Test. *Management Science*, *34*(2), 139–159. https://doi.org/10.1287/mnsc.34.2.139

Sharma, B., Gao, Y., Miller, T., Churpek, M. M., Afshar, M., & Dligach, D. (2023). *Multi-Task Training with In-Domain Language Models for Diagnostic Reasoning* (arXiv:2306.04551). arXiv. https://doi.org/10.48550/arXiv.2306.04551

Shortliffe, E. H. (1974). A rule-based computer program for advising physicians regarding antimicrobial therapy selection. *Proceedings of the 1974 Annual ACM Conference - Volume 2*, 739. https://doi.org/10.1145/1408800.1408906

Singh, A., Ehtesham, A., Kumar, S., & Khoei, T. T. (2024). Enhancing AI Systems with Agentic Workflows Patterns in Large Language Model. *2024 IEEE World AI IoT Congress (AIIoT)*, 527–532. https://doi.org/10.1109/AIIoT61789.2024.10578990

Sowa, K., & Przegalinska, A. (2020). Digital Coworker: Human-AI Collaboration in Work Environment, on the Example of Virtual Assistants for Management Professions. In A. Przegalinska, F. Grippa, & P. A. Gloor (Eds.), *Digital Transformation of Collaboration* (pp. 179–201). Springer International Publishing. https://doi.org/10.1007/978-3-030-48993-9_13

Sowa, K., Przegalinska, A., & Ciechanowski, L. (2021). Cobots in knowledge work: Human – AI collaboration in managerial professions. *Journal of Business Research*, *125*, 135–142. https://doi.org/10.1016/j.jbusres.2020.11.038

Spitzer, P., Kühl, N., & Goutier, M. (2022). *Training Novices: The Role of Human-AI Collaboration and Knowledge Transfer* (arXiv:2207.00497). arXiv. https://doi.org/10.48550/arXiv.2207.00497

Suzuki, M., & Matsuo, Y. (2022). A survey of multimodal deep generative models. *Advanced Robotics*, *36*(5–6), 261–278. https://doi.org/10.1080/01691864.2022.2035253

Terren, L. T. L., & Borge-Bravo, R. B.-B. R. (2021). Echo Chambers on Social Media: A Systematic Review of the Literature. *Review of Communication Research*, *9*. https://doi.org/10.12840/ISSN.2255-4165.028

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis. *Current Anthropology*, *53*(6), 673–692. https://doi.org/10.1086/668207

Tschang, F. T., & Almirall, E. (2021). Artificial Intelligence as Augmenting Automation: Implications for Employment. *Academy of Management Perspectives*, *35*(4), 642–659. https://doi.org/10.5465/amp.2019.0062

Volz, K. M., & Dorneich, M. C. (2020). Evaluation of Cognitive Skill Degradation in Flight Planning. *Journal of Cognitive Engineering and Decision Making*, *14*(4), 263–287. https://doi.org/10.1177/1555343420962897

Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020). From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3334480.3381069

Wang, F., Liu, X., Liu, O., Neshati, A., Ma, T., Zhu, M., & Zhao, J. (2023). Slide4N: Creating Presentation Slides from Computational Notebooks with Human-AI Collaboration. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3544548.3580753

Wang, Z., Valdez, J., Basu Mallick, D., & Baraniuk, R. G. (2022). Towards Human-Like Educational Question Generation with Large Language Models. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 153–166). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_13

Waterman, D. A. (1985). *A guide to expert systems*. Addison-Wesley Longman Publishing Co., Inc.

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy. *Work, Employment and Society*, *33*(1), 56–75. https://doi.org/10.1177/0950017018785616

Wu, J., Thorne-Large, J., & Zhang, P. (2022). Safety first: The risk of over-reliance on technology in navigation. *Journal of Transportation Safety & Security*, *14*(7), 1220–1246. https://doi.org/10.1080/19439962.2021.1909681

Yetistiren, B., Ozsoy, I., & Tuzun, E. (2022). Assessing the quality of GitHub copilot's code generation. *Proceedings of the 18th International Conference on Predictive Models and Data Analytics in Software Engineering*, 62–71. https://doi.org/10.1145/3558489.3559072

Yoon, Y., Guimaraes, T., & Swales, G. (1994). Integrating artificial neural networks with rule-based expert systems. *Decision Support Systems*, *11*(5), 497–507. https://doi.org/10.1016/0167-9236(94)90021-3