



CNaPPES.23

9º Congresso Nacional de Práticas Pedagógicas
no Ensino Superior

Livro de Atas

**9.º Congresso Nacional
de Práticas Pedagógicas
no Ensino Superior**

6 e 7 de julho de 2023

FICHA TÉCNICA

Título

Livro de Atas do 9.º Congresso Nacional
de Práticas Pedagógicas no Ensino Superior

Coordenação

Eduardo Esteves
Dulce Estêvão
Jânio Monteiro
Marisol Correia

Design

CO CNaPPES-23

Editora

Universidade do Algarve

1.ª edição – maio de 2024

ISBN

978-989-9127-67-8

DOI

<https://doi.org/10.34623/kf51-wf33>

Handle do Sapiëntia

<http://hdl.handle.net/10400.1/20248>

Os conteúdos apresentados são da exclusiva responsabilidade dos respetivos autores.
© Autores. Esta obra encontra-se sob a Licença Internacional Creative Commons Atribuição 4.0

Inteligência Artificial na avaliação tradicional: aquisição de conhecimento vs Prompt Engineering

João M. C. Estêvão ^{1,2}

M. Dulce Estêvão ³

¹ Instituto Superior de Engenharia, Universidade do Algarve

² CIMA-UAlg
jestevao@ualg.pt

² Escola Superior de Saúde, Universidade do Algarve

mestevao@ualg.pt

Resumo

No atual contexto do desenvolvimento da inteligência artificial (IA), e da sua ampla divulgação pelo público em geral, este trabalho pretendeu testar as capacidades das atuais ferramentas de IA na resolução de testes escritos, e perceber quais as competências a adquirir neste contexto de avanço tecnológico. Para isso, foram resolvidos três exames de uma unidade curricular (UC) de Bioquímica, de forma sequencial, por um utilizador do ChatGPT, com experiência em IA, mas sem conhecimentos de Bioquímica. As provas foram corrigidas pela docente da UC, tal como corrigidas para os alunos, tendo as correções sido disponibilizadas ao utilizador do ChatGPT, antes da realização da prova seguinte. Desta forma, foi possível melhorar as instruções utilizadas no ChatGPT, para melhorar a qualidade das respostas. Os resultados mostram que foi possível obter classificações superiores às classificações obtidas pelos alunos, apenas com o recurso ao ChatGPT e sem qualquer conhecimento sobre as matérias avaliadas. Pode concluir-se que a qualidade das instruções fornecidas (*prompts*) são um dos fatores determinantes para a obtenção de respostas mais corretas quando se utiliza este tipo de ferramentas e que poderá ser o momento para equacionar a integração progressiva e controlada das tecnologias da IA nos novos modelos de formação.

Palavras-Chave: Inteligência artificial, *Large Language Models*, *prompt engineering*, testes escritos.

1. Contextualização

O relatório do Fórum Económico Mundial sobre a evolução dos empregos entre 2023-2027 (WEF, 2023) indica que a Inteligência Artificial (IA) poderá ter um efeito disruptivo no mercado de trabalho, resultando na eliminação de alguns tipos de emprego, criação de novos, e alteração das competências necessárias para as profissões que persistirem. Este tipo de relatório propicia o surgimento, na sociedade, do tipo de imagem da IA que os filmes de ficção científica costumam apresentar. O filme “2001 Odisseia no espaço” (1968), realizado por Stanley Kubrick, criado pelo realizador em conjunto com Arthur C. Clarke, é um dos primeiros filmes onde este tema é abordado e onde é transmitida a ideia de que a IA não erra, mas que pode tornar-se maléfica e um risco para a Humanidade. Contudo, este tipo de mensagem não será a mais correta, à luz do atual estado do conhecimento,

dado que a IA ainda erra muitas vezes, e que não é provável que venha a ter o efeito “maléfico” de acabar com o emprego, de forma generalizada, eliminando totalmente as pessoas na realização das suas tarefas. Num futuro próximo, o que poderemos esperar será uma substituição das pessoas que não usam a IA por pessoas capazes de a utilizar, e não pela própria IA. O número de utilizadores da IA poderá até superar as atuais necessidades de trabalhadores, mas estes precisarão de novas competências e conhecimentos, o que poderá gerar um aumento de produtividade. Neste contexto, a Sociedade em geral, mas, muito especificamente, as instituições do ensino devem estar atentas a esta realidade e adaptar-se, tornando-se necessária uma profunda análise e debate em torno deste tema (Gimpel et al., 2023; Pelletier et al., 2023).

1.1. Introdução à Inteligência Artificial (IA)

Antes de se utilizar a IA, é desejável adquirirmos alguns conhecimentos básicos sobre o assunto. As bases da IA tiveram o seu início na década de 1950s (Delipetrev et al., 2020). Com o desenvolvimento das redes neuronais artificiais, a IA ganhou um novo ímpeto nos anos 1990s, tendo sido utilizada com sucesso em numerosas aplicações, normalmente sem que os utilizadores tivessem consciência da sua utilização. Foi com o desenvolvimento dos modelos de linguagem de grande escala LLM (*Large Language Models*), e, mais recentemente, com o lançamento do ChatGPT, em novembro de 2022, que o grande público começou a ter consciência das atuais capacidades da IA.

1.2. Modelos de linguagem generativa

Os modelos LLM são modelos de IA previamente treinados com grandes quantidades de textos para aprenderem a gerar linguagem humana (Ray, 2023). O ChatGPT é um exemplo de um LLM que usa a tecnologia GPT (*Generative Pre-trained Transformer*). Estes modelos generativos de linguagem foram desenvolvidos para gerarem texto, mas têm sido observadas outras capacidades, que parecem imitar o raciocínio humano, o que os tornou populares muito rapidamente. Atualmente, o ChatGPT (OpenAI) adota o GPT-3.5 na versão gratuita, e o GPT-4 na versão que requer subscrição. O chat do Bing (Microsoft) também usa o GPT-4, mas é atualmente gratuito. Por sua vez, o Bard (Google), também gratuito (até à data), baseia-se num modelo distinto, designado PaLM, que possui uma arquitetura diferente (Chowdhery et al., 2022).

Os LLM são constituídos por sistemas de aprendizagem automática (*machine learning*), tais como as redes neuronais (classificadas como sistemas de *deep learning*), de modo a gerarem respostas semelhantes às dos humanos, num elevado número de línguas. Estes modelos foram sujeitos a uma primeira fase de treino em larga escala, não supervisionado, usando a informação disponível na internet, como livros, artigos, páginas da internet, entre outros. Normalmente, estes textos são primeiro transformados em unidades fundamentais de texto (os *tokens*), que poderão corresponder a palavras inteiras, ou somente a parte de palavras, que depois são convertidos em números, alimentando os sistemas de aprendizagem automática profunda (*deep learning*), com muitos milhões de ligações. Os LLM fazem a previsão do próximo *token* com base na sequência de *tokens* existentes nos *prompts* (as instruções fornecidas aos LLM), que são criados pelo utilizador (Beurer-Kellner et al., 2023). A *prompt engineering* é uma nova área do conhecimento que visa o desenvolvimento e a otimização de *prompts* a usar, de forma eficaz, nos LLM (Lo, White et al., 2023). Alguns estudos têm demonstrado que os diversos LLM atualmente existentes apresentam sensibilidades diferentes aos *prompts* (López Espejel et al., 2023).

1.3. Prompt Engineering

Um *prompt* é um conjunto de instruções que fornecemos a um LLM de modo a obter um resultado. Os *prompts* podem influenciar muito os resultados gerados pelos LLM, incluindo

o rigor desses resultados. Neste contexto, a *prompt engineering* começa a ganhar uma especial importância (White et al., 2023), na tentativa de minimizar o número de alucinações dos LLM.

No contexto dos LLM, “alucinação” é o termo usualmente empregue para designar uma resposta incorreta e/ou indesejável (Ji et al., 2023). As alucinações podem 1) ocorrer ao nível da coerência da resposta (na mesma resposta, o LLM pode apresentar uma versão e o seu contrário); 2) contradizer o *prompt* apresentado (por exemplo, a pergunta é do tipo afirmativo, mas o texto da resposta é do tipo negativo); 3) contradizer os factos (designadamente gerando uma resposta incorreta do ponto de vista factual); ou 4) apresentar informação fora de contexto (que até pode ser correta, mas que não está enquadrada no conjunto de instruções fornecidas ao LLM).

Existem diversos motivos possíveis para as alucinações dos LLM ocorrerem. O motivo mais evidente está relacionado com a qualidade dos dados usados no treino dos LLM. Se não existem dados sobre um determinado assunto nas fontes de informação usadas para o treino de um LLM, então esse LLM terá tendência para alucinar sobre esse assunto, eventualmente apresentando um texto coerente do ponto de vista gramatical, mas factualmente incorreto. Assim, atualmente, a probabilidade de obtermos uma alucinação é tanto maior quanto mais especializado for o assunto, nomeadamente se o assunto em causa não estiver corretamente apresentado em páginas da internet de acesso livre. Uma outra causa de alucinações está relacionada com o tipo de LLM. A taxa de alucinação obtida com o GPT-3.5 é superior à do GPT-4, e é diferente da que se obtém com o PaLM 2, por exemplo (López Espejel et al., 2023). Uma outra fonte de alucinações, está relacionada com a qualidade dos *prompts* que são utilizados. Um *prompt* bem estruturado pode orientar o LLM para a obtenção de um resultado correto, enquanto um “mau” *prompt* pode originar uma alucinação (White et al., 2023). Esta é a fonte de alucinações que um utilizador de um LLM pode controlar mais facilmente.

É neste contexto que emerge a necessidade de desenvolvimento da *prompt engineering*, para o qual as capacidades humanas, tais como a criatividade e a intuição, têm um papel importante, para além do conhecimento interdisciplinar (Lo, 2023).

Um “bom” *prompt* deve ser conciso (com instruções breves e claras), lógico (com instruções estruturadas e coerentes), explícito (com instruções explícitas sobre a forma de apresentar os resultados), adaptativo (com instruções equilibradas entre a criatividade e a especificidade) e refletivo (a melhoria dos *prompts* deve resultar da realização de múltiplos testes que avaliem a sua eficácia em assuntos conhecidos, nomeadamente relativamente ao rigor, coerência e utilidade das respostas no contexto das instruções criadas) (Lo, 2023). Por outro lado, também podemos usar algumas estratégias ativas adicionais para minimizar os problemas relacionados com as alucinações. Uma dessas estratégias passa pelo ajuste da “criatividade” (*temperature*) da resposta, quando essa opção existe. Isto significa que pode ser feito um ajuste do critério de variabilidade da seleção das respostas dadas pelo LLM, em função da probabilidade de obtenção de um determinado resultado. Esta probabilidade pode ser condicionada pelo conjunto de instruções fornecidas ao LLM, sendo que a escolha de uma maior criatividade poderá originar uma maior probabilidade do modelo alucinar, mas, por outro lado, poderá fornecer “ideias” mais inovadoras para o nosso problema/questão.

O chamado *zero-shot prompt* é o tipo mais básico de instrução, no qual não se fornece nenhum contexto sobre o assunto que se está a pesquisar, mas se espera que o LLM forneça uma resposta adequada usando o seu conhecimento geral (por exemplo, “Descreve o processo de fotossíntese.”).

Uma forma de melhorar os resultados dos LLM é recorrer aos designados *multiple-shot prompts*. Neste tipo de *prompts* são apresentadas diversas instruções, de modo a fornecer ao LLM o contexto necessário para reconhecer padrões, através da apresentação de uma sequência lógica de exemplos e instruções, o que reduz a probabilidade do modelo alucinar (Brown et al., 2020). Os *one-shot prompts* são um caso particular, em que só é fornecido um exemplo para que o LLM possa replicar esse padrão corretamente num outro problema (por exemplo: “Um prato tinha 3 maçãs. A Ana comeu uma e o Joaquim colocou no prato mais duas maçãs. Quantas maçãs ficaram no prato? Resposta: $3-1+2 = 4$ maçãs. Com base

neste exemplo, resolve o seguinte problema: ...”). Este tipo de *prompts* exige algum conhecimento sobre o assunto, para que seja apresentado ao LLM um exemplo correto. Também estão a surgir outras estratégias mais elaboradas, como, por exemplo o recurso a árvores de pensamento (Yao et al., 2023), onde as instruções são dadas ao LLM de modo que este crie diferentes linhas de pensamento (podem ser diferentes entidades, tais como pessoas virtuais, por exemplo). Estas estratégias podem levar a uma solução mais consensual, que poderá ser mais exata.

Assim, parece que a combinação da criatividade do utilizador na elaboração dos *prompts*, com o seu domínio em *prompt engineering* (engenharia de *prompts*), constitui um elemento essencial que pode influenciar a precisão dos resultados dos LLM, sendo crucial testar e avaliar esta dinâmica no contexto do ensino superior.

2. Descrição da prática pedagógica

Tendo em conta o atual desenvolvimento das capacidades dos LLM, surgiram as seguintes perguntas: Será possível conseguir uma aprovação numa prova “tradicional” de avaliação escrita de conhecimentos, ao nível do ensino superior, por alguém que não tenha qualquer conhecimento específico relativo a uma determinada matéria, somente usando este tipo de aplicações informáticas? Se sim, quais são os conhecimentos e/ou as competências a transmitir no futuro, no ensino superior, tendo em conta a previsível evolução do mercado de trabalho? Este estudo foi realizado neste contexto.

2.1. Objetivos e público-alvo

O objetivo deste estudo foi demonstrar as capacidades atuais dos modelos LLM na resolução de provas de avaliação tradicionais (teste escrito), e perceber quais as competências que devem ser adquiridas neste novo contexto de avanço tecnológico.

2.2. Metodologia

De modo a avaliar se, no atual contexto de desenvolvimento dos LLM, é possível obter aprovação numa prova tradicional de avaliação escrita, por alguém que não possui conhecimento específico sobre um determinado assunto, foi adotada a seguinte metodologia (Figura 1):

- Foram resolvidos três exames da unidade curricular (UC) de Bioquímica do curso de Farmácia da Escola Superior de Saúde da Universidade do Algarve (ESSUAlg), de forma sequencial, por um Engenheiro Civil, que nunca estudou o assunto, mas que possui conhecimentos e vasta experiência na programação de *software* com algoritmos de *machine learning*, de uma forma geral, e de algoritmos de *deep learning*, em particular (redes neuronais artificiais);
- Foi usado o ChatGPT (com o GPT-3.5) para resolver as questões, sem recorrer a um outro tipo de consulta (internet ou livros), apresentando os *prompts* sempre em português Europeu (embora tenham sido feitos, posteriormente, alguns testes que permitiram constatar alguma melhoria na qualidade das respostas quando se recorre ao inglês, por exemplo, pois existem mais fontes de informação nessa língua);
- A docente responsável pela referida UC corrigiu as resoluções dos exames realizados com respostas fornecidas pelo ChatGPT, do mesmo modo que corrigiu as respostas dos alunos da UC que foram avaliados por exame, tendo transmitido a nota e facultado a correção após a resolução de cada exame, ao utilizador do ChatGPT;
- As notas obtidas com o ChatGPT foram depois comparadas com as notas obtidas pelos alunos que realizaram cada uma das provas (é de salientar que parte dos

alunos já tinham dispensado de exame), de modo a ser avaliada a capacidade dos LLM em fornecer respostas ao nível do ensino superior, e a importância relativa da *prompt engineering*.

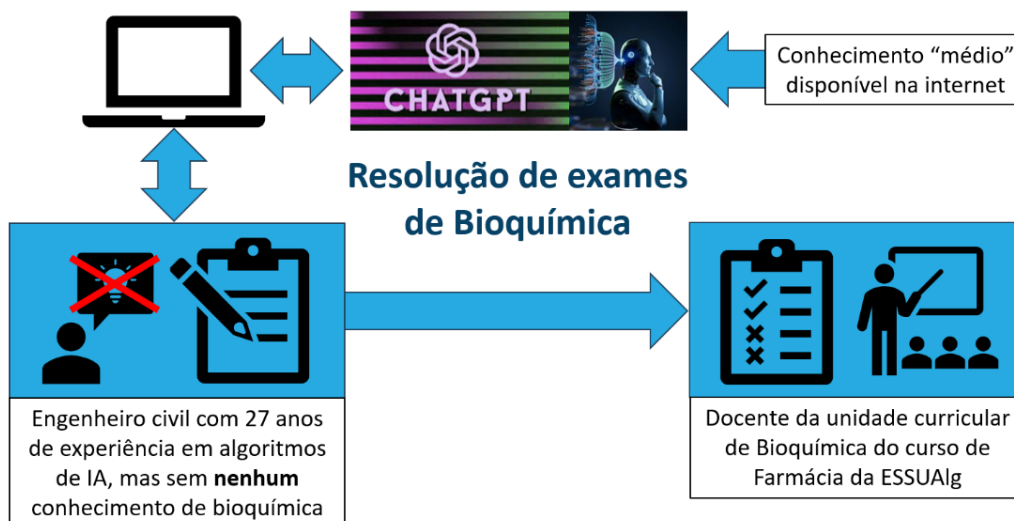


Figura 1. Esquema que resume a experiência pedagógica realizada.

2.3. Avaliação

Depois de resolvido cada exame de Bioquímica, com o recurso ao ChatGPT, por um utilizador que não possuía qualquer tipo de conhecimento específico sobre o assunto, foi feita uma análise detalhada aos erros cometidos, de modo a ser possível perceber quais os tipos de *prompts* que originavam mais erros nas respostas (com diversos tipos de alucinações), e os que permitiam obter uma resposta correta.

Esta estratégia iterativa permitiu melhorar o desenvolvimento de *prompts* eficazes, o que possibilitou uma avaliação da importância da *prompt engineering* no contexto da utilização do ChatGPT, designadamente na resolução de provas tradicionais de avaliação escrita.

As notas obtidas através do uso do ChatGPT foram comparadas com os valores médios obtidos pelos alunos que realizaram cada um dos exames, e com a classificação mais elevada obtida.

Foram escolhidos exames de modo que a sequência apresentasse uma diminuição progressiva dos resultados dos alunos, tanto em termos médios, como em relação às melhores classificações, para que os resultados obtidos com o ChatGPT evidenciasse a importância da *prompt engineering* neste contexto, já que esta era a hipótese em estudo.

3. Resultados, implicações e recomendações

Os resultados obtidos (Tabela 1) permitem concluir que foi possível obter aprovação em todos os exames realizados, somente com o recurso ao ChatGPT (com classificações entre 11,05 e 17,21 valores, numa escala até vinte valores). Os resultados obtidos com o ChatGPT foram sempre superiores à média obtida pelos alunos, e no segundo e terceiro exames foram muito superiores à melhor classificação obtida pelos alunos.

Quando analisamos em detalhe os resultados obtidos, é possível constatar que a classificação no segundo e no terceiro exame foi muito superior à classificação obtida no primeiro. Essa subida não resultou do estudo e aquisição de conhecimentos sobre as matérias lecionadas na UC de Bioquímica em avaliação, mas sim do aumento do conhecimento sobre a *prompt engineering*, por parte de quem utilizou o ChatGPT. Essa

melhoria das competências resultou da leitura de diversos artigos de especialidade no intervalo de tempo decorrido entre a resolução do primeiro exame e os restantes, e da análise aos erros cometidos pelo ChatGPT na resolução do primeiro exame (devido a uma menor eficácia dos *prompts* usados nessa primeira prova de avaliação). Ou seja, a melhoria dos resultados deveu-se apenas à melhoria das competências adquiridas sobre a utilização eficaz dos LLM e não à aquisição de conhecimentos sobre as matérias avaliadas.

Tabela 1. Classificações obtidas nos exames de Bioquímica.

Anos letivos dos exames*	Alunos		ChatGPT
	Melhor resultado	Média (nº alunos)	
1) 2022/23	14,25	7,27 (25)	11,05
2) 2019/20	12,48	12,48 (1)	17,21
3) 2016/17	8,10	4,45 (4)	16,40

* Pela ordem de resolução.

Uma questão que ficou evidente com este trabalho foi a importância da criatividade na redação de *prompts* eficazes. Por exemplo, como procurar a resposta correta para uma pergunta em que era necessário completar espaços vazios numa frase: “A lipogénese que ocorre no tecido adiposo pode ser _____ por ação da insulina, aumentando assim o armazenamento de _____ nos adipócitos.”?

Para ilustrar o tipo de estratégias que podem ser adotadas, poderíamos recorrer ao seguinte *prompt* para resolver a questão anterior:

- Considera que és um professor de Bioquímica do ensino superior.
- Explica o que é a lipogénese.
- Explica o que são adipócitos.
- Lista os principais tipos de ação da insulina no corpo humano.
- Substitui AAA e BBB de modo a completar corretamente a seguinte frase: "A lipogénese que ocorre no tecido adiposo pode ser AAA por ação da insulina, aumentando assim o armazenamento de BBB nos adipócitos."

A primeira instrução serve para estabelecer o nível de rigor e a profundidade na apresentação da resposta.

As três instruções seguintes dão mais contexto ao LLM sobre os assuntos em questão, e permitem ao utilizador uma aferição da coerência da resposta, especialmente para quem não domina as matérias em causa (pois possibilita a identificação de algum grau de alucinação, por falta de coerência entre as definições apresentadas e a resposta final).

A utilização de AAA e BBB, em vez dos espaços vazios, é uma estratégia que permite transmitir ao LLM o objetivo pretendido.

Contudo, ainda assim poderemos obter respostas diferentes quando usamos diferentes LLM (por exemplo o ChatGPT, o chat do Bing, ou o Bard), ou quando alteramos a “criatividade” do modelo (no chat do Bing, por exemplo, escolhendo um dos três modos atualmente disponíveis: preciso, equilibrado ou criativo). Além disso, como os modelos são probabilísticos, não é garantido que a resposta será sempre a mesma em todas as tentativas de obtenção de resposta, mesmo quando é utilizado o mesmo *prompt*, e que não existe algum grau de alucinação, ainda que possa ser ligeiro. Esta variabilidade deve ser sempre tida em consideração quando se usam os LLM, para qualquer finalidade, devendo, portanto, haver sempre muito cuidado na interpretação e aceitação dos resultados obtidos. Os resultados obtidos neste trabalho demonstram que é possível atingir um desempenho muito superior ao do “aluno médio”, se minimizarmos a taxa de alucinação (que nunca será nula), com o recurso apropriado à *prompt engineering*.

4. Conclusões

Os resultados obtidos com a metodologia adotada permitiram demonstrar que é possível obter aprovação numa UC do ensino superior sem ter qualquer tipo de conhecimento sobre

o assunto em avaliação, através da resolução de uma prova escrita tradicional de avaliação de conhecimentos, somente recorrendo às respostas fornecidas pelo ChatGPT inclusivamente com resultados superiores às classificações médias obtidas pelos alunos em cada uma das provas testadas.

As instruções (*prompts*) são determinantes para a obtenção de respostas corretas, pelo que estas ferramentas (LLM) devem ser utilizadas com precaução, e com consciência das suas limitações e vantagens uma vez que nem sempre os resultados são fiáveis, embora a combinação do conhecimento especializado com estas tecnologias poderá facilitar a obtenção de soluções mais rápidas e precisas.

Parece ser desejável uma integração progressiva e controlada das tecnologias da IA nos novos modelos de formação, com a orientação dos docentes relativamente à melhor forma de utilização e com a adoção de modelos alternativos de avaliação que minimizem a utilização abusiva e inadequada dos LLM para a obtenção de resultados académicos positivos.

5. Referências Bibliográficas

- Beurer-Kellner, L., Fischer, M., & Vechev, M. (2023). Prompting Is Programming: A Query Language for Large Language Models. *Proceedings of the ACM on Programming Languages*, 7(Issue PLDI, Article No.: 186), 1946–1969. <https://doi.org/10.5281/zenodo.7711823>
- 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., ... Amodei, D. (2020). *Language models are few-shot learners* Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada.
- Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N. M., Prabhakaran, V., ... Fiedel, N. (2022). PaLM: Scaling Language Modeling with Pathways. *Journal of Machine Learning Research*, 24, 240:241-240:113.
- Delipetrev, B., Tsinarakii, C., & Kostić, U. (2020). *Historical Evolution of Artificial Intelligence* (EUR 30221 EN). Publications Office of the European Union.
- Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., Röglinger, M., Ruiner, C., Schoch, M., Schoop, M., Urbach, N., & Vandirck, S. (2023). *Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers*. University of Hohenheim.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>
- Lo, L. S. (2023). The Art and Science of Prompt Engineering: A New Literacy in the Information Age. *Internet Reference Services Quarterly*, 1-8. <https://doi.org/10.1080/10875301.2023.2227621>
- López Espejel, J., Ettifouri, E. H., Yahaya Alassan, M. S., Chouham, E. M., & Dahhane, W. (2023). GPT-3.5, GPT-4, or BARD? Evaluating LLMs reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal*, 5, 100032. <https://doi.org/10.1016/j.nlp.2023.100032>
- Pelletier, K., Robert, J., Muscanell, N., McCormack, M., Reeves, J., Arbino, N., Grajek, S., Birdwell, T., Liu, D., Mandernach, J., Moore, A., Porcaro, A., Rutledge, R., & Zimmern, J. (2023). *2023 EDUCAUSE Horizon Report, Teaching and Learning Edition*. EDUCAUSE.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- WEF. (2023). *Future of Jobs Report*. World Economic Forum.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *ArXiv, abs/2302.11382*.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). Tree of Thoughts: Deliberate Problem Solving with Large Language Models. *ArXiv*, *abs/2305.10601*.