Gen Re®

Generative AI

September 11, 2023

This presentation has been compiled for educational purposes.

Table of Contents

Resources

- An accessible article on deep learning is the 2015 *Nature* paper "Deep learning" by Yann LeCun, Yoshua Bengio, Geoffrey Hinton.[1] The three authors are among the most influential contributors to the development of deep learning methods.[2]
- An accessible course on deep learning is available from Melissa Dell at Harvard (*Unleashing Novel Data at Scale*, Spring 2023).[3]
- An excellent introduction into neural networks is the three-part *3Blue1Brown* video series (recommended by Melissa Dell in the course referenced above) available at

  https://www.youtube.com/watch?v=aircAruvnKk
  https://www.youtube.com/watch?v=IHZwWFHWa-w
  https://www.youtube.com/watch?v=Ilg3gGewQ5U.

- For best practices, see OpenAI, https://platform.openai.com/docs/guides/gpt-best-practices.[4]

1) The paper is available from Geoffrey Hinton's webpage at the University of Toronto: https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf.
2) In the academic field of artificial intelligence, the most senior author is listed last.
3) Melissa Dell, *Knowledge Base*, https://dell-research-harvard.github.io/blog.html.
4) Best practices for OpenAI's ChatGPT may not serve as best practices for Google's Bard.

Large Language Models

Large Language Models

- A Large Language Model (LLM) is a *deep learning* algorithm that can recognize, summarize, translate, predict, and generate text and other content.
- LLMs are typically trained on massive datasets, large enough to include nearly everything that has been written on the public Web over a large span of time.[1,2]
- Using *unsupervised (aka self-supervised) learning*, the AI algorithm learns words as well as the relationships between and concepts behind them.[3]
- Such *pre-trained* models are subject to ongoing *reinforcement learning by human feedback* (RLHF) following deployment.[4]
- Large language models are unlocking possibilities in areas such as search engines, natural language processing, healthcare, robotics, and code generation.

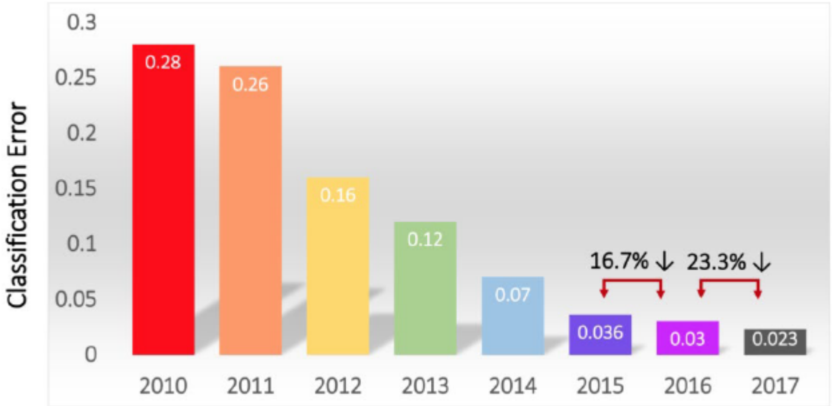
1) Developers of Large Language Models (such as Google, OpenAI, or Meta) may augment publicly available data with licensed data and with synthetic data.
2) GPT-4, which is the latest OpenAI LLM, was trained on data, the vast majority of which cuts off in September 2021. "...GPT-4 does not learn from its experience." https://openai.com/research/gpt-4.
3) In unsupervised learning, the algorithm makes sense of unlabeled data on its own by performing feature detection and classification. LLMs learn predicting the next work (GPT) or masked tokens (BERT).
4) RLHF is a type of reinforcement learning in which an AI system receives feedback from human trainers to improve the decisions it makes.
Source: Nvidia, "What Are Large Language Models Used For?" January 26, 2023, https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-models-used-for/.

The Deep Learning Revolution (1/2)

"From not working to neural networking" *(The Economist)*[1]

- The Deep Learning revolution was unleashed in 2012, when AlexNet, a convolutional neural network (CNN) running on Graphics Processing Units (GPUs) won the ImageNet contest by a wide margin.[2,3]
- Approaches based on feature engineering rely on experts manually designing hand-crafted features based on their domain knowledge.
- Deep learning, on the other hand, integrates both feature learning and classification.

ImageNet Winners



[Feature Engineering] [Convolutional Neural Networks →]

The ranking is based on the top-5 classification error, which is the percentage of times that the target label does not appear among the 5 highest-probability predictions.
Source: Kaggle, https://www.kaggle.com/discussions/getting-started/149448.

1) *The Economist* (June 23, 2016) "From not working to neural networking." https://www.economist.com/special-report/2016/06/23/from-not-working-to-neural-networking.
2) The ImageNet competition was about predicting the predetermined class of each image. The competition has been discontinued after the error rate of the neural networks dropped below the human error rate.
3) AlexNet was designed by Alex Krizhevsky in collaboration with Ilya Sutskever and his Ph.D. advisor Geoffrey Hinton. Although not the first CNN on GPUs, the margin by which it beat its competitors was stunning. For the technical paper, see NVIDIA, https://www.nvidia.cn/content/tesla/pdf/machine-learning/imagenet-classification-with-deep-convolutional-nn.pdf.

The Deep Learning Revolution (2/2)

- The following four factors are credited with revolutionizing neural networks:[1]
  1. The availability of vast amounts of data for training (harvested from the public Web).
     - ImageNet offered a vast trove of labeled images to train classification models on.
  2. The computing power of graphics processing units.
     - NVIDIA chips and its CUDA[2] software layer.
  3. A new approach to initializing weights in the layers of the neural network.[3]
  4. Change of the activation function from the sigmoid to the partially linear ReLU.[4,5]

1) See Melissa Dell, *Unleashing Novel Data at Scale*, Spring 2023, Lecture 2, https://dell-research-harvard.github.io/intro/2023/01/19/lecture2.html.
2) Compute Unified Device Architecture.
3) See Yann LeCun, Yoshua Bengio, and Geoffrey Hinton (2015) "Deep learning," Nature. https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf.
4) Rectified linear unit.
5) See LeCun, Bengio, and Hinton, *op. cit.*

The Transformer

"Attention is All You Need"

- In 2017, Google introduced the Transformer.
- Now the dominant neural network architecture, the model is based on a self-attention mechanism, which directly models relationships among all words in a sentence.[1,2]
  - For example, deciding on the most likely meaning of the word "bank" in the sentence "I arrived at the bank after crossing the…" requires knowing if the sentence ends in "… road." or "… river."
  - To determine that the word "bank" refers to the shore of a river and not a firm, the Transformer can learn to attend immediately to the word "river" and make this decision in a single step.
  - In comparison, a recursive neural network (RNN) could only determine that the word "bank" is likely to refer to the bank of a river after reading each word between "bank" and "river" step by step.[3]
  - Although CNNs are much less sequential than RNNs, the number of steps it takes CNNs to combine information from distant parts of the input still grows with increasing distance.

1) See the Google blog post by Jakob Uszkoreit, "Transformer: A Novel Neural Network Architecture for Language Understanding," https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html.
2) GPT is a transformer. GPT-4-32-k, released in May 2023, has a context window of 32,768 tokens (which translates into 24,000 words, approximately). See OpenAI, https://platform.openai.com/docs/models/gpt-4.
2) At the time the Transformer was created, RNNs were at the core of the leading approaches to language understanding tasks such as language modeling, machine translation, and question answering.

Text Representation and Text Generation (1/3)

- Semantic Search
  - Semantic search seeks to improve search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable dataspace, whether on the Web or within a closed system, to generate more relevant results.
    - Microsoft's Bing Chat (Enterprise)—sidebar in Edge browser.
    - Google SGE (Search Generative Experience) and Google Bard—requires signing into Google account.
- Summarization
  - Log into ChatGPT or Google Bard and paste the following *prompt* into the dialog box:

    Please summarize in one paragraph: [insert text of Hansel and Gretel from https://sites.pitt.edu/~dash/grimm015.html]

    or

    Please summarize in a poem: […]

Some of the above is sourced from OpenAI, https://platform.openai.com/.

Text Representation and Text Generation (2/3)

- Synthetization
  - Data synthesis brings together results and examines the findings together for patterns of agreement, convergence, divergence, or discrepancy.
    - Log into ChatGPT or Google Bard and paste the following *prompt* into the dialog box:

      My cat weighs 20 pounds.
    - In the *completion*, the LLM will compare the cat's weight to what is considered healthy.
- Classification, categorization, and sentiment analysis
  - Classification assigns objects to predefined groups.
  - Categorization groups objects based on shared characteristics.
  - Sentiment analysis is the process of analyzing text to determine, for instance, if the emotional tone of a message is positive, negative, or neutral.

Some of the above is sourced from OpenAI, https://platform.openai.com/.

Text Representation and Text Generation (3/3)

- Translation
  - Transformer models were developed in the context of translation.
  - OpenAI's GPT-4 claims high accuracy for a large set of languages, including low-resource languages.[1]
- Data extraction
  - Data extraction from documents comprises image analysis and optical character recognition (OCR).[2]
  - OCR typically uses a vision model for feature extraction and a language model for text transcription.[3]
- Content creation
  - OpenAI's ChatGPT and Google's Bard are tools for creating text documents on desired topics, of desired length, and of desired form (blog post, essay, poem, etc.).
  - OpenAI's DALL-E[4] and Google's Imagen generate images from text (which is processed using an LLM).

1) OpenAI, GPT-4 Technical Report, March 27, 2023, https://arxiv.org/pdf/2303.08774.pdf.
2) Melissa Dell published Layout Parser, an open-source deep-learning powered library for processing document image data at scale. See https://dell-research-harvard.github.io/resources/layout-parser.
3) The language model is based on a seq2seq ([text] sequence to [text] sequence) architecture—examples are EasyOCR and TrOCR. For a vision-only OCR architecture, see SVTR and Efficienct OCR (EffOCR).
4) The name DALL-E is held to be a word play on Salvador Dali and Pixar's WALL-E.
Some of the above is sourced from OpenAI, https://platform.openai.com/.

Generative Search

- OpenAI's ChatGPT and Google's Bard are AI tools for representing and generating text and as such do not offer information on sources of information.
  - The corpus the LLM was pre-trained on has a cut-off, which is September 2021 for the latest version of GPT and mid-2021 for Bard's LLM LaMDA.
- In *generative search*, on the other hand, an LLM is augmented by a *retriever model* (RM).[1]
  - The LLM breaks down the prompt into a chain of thoughts. The RM retrieves knowledge from an external, retrievable data base (e.g., the Web). The LLM then summarizes the retrieved knowledge, engages the user in a conversation, and generates content (e.g., a corporate memo).
- Prominent *generative search engines* are Microsoft's Bing Chat and Google's SGE.
- Bard also augments its LLM with an RM, providing real-time information from the web.[2,3]

1) The architecture is known as RAG (retriever-augmented generation). A retriever-augmented LLM is known as RALM (retriever-augmented language model).

2) Although Bard retrieves real-time information from the web, it does not provide sources, unless it quotes at length from a webpage or returns an image. See Google Bard FAQ "How and when does Bard cite sources in its responses?", https://bard.google.com/faq#citation.

3) See also "Benefits of Generative Search: Unlocking Real-Time Knowledge Access," Gen Re, August 17, 2023, https://www.genre.com/us/knowledge/publications/2023/august/benefits-of-generative-search-unlocking-real-time-knowledge-access-en.

Engineering Tools (1/3)

- *Prompt design* – General business user
  - Six best practices of prompt design:
    - Include details in your query to get more relevant answers.
    - Ask the model to adopt a *persona*, of which the following is an example:

      You are an AI assistant who has been working in the insurance industry for years. You have a deep understanding of business etiquette and are fluent in the jargon used in my field, which is reinsurance. Your tone is formal and respectful, and you always provide well-reasoned and data-driven advice.

      Next, ask a question within the same chat (which has conversational memory):

      Please draft an email for me to send to a client about a late bordereau.
    - Use delimiters to clearly indicate distinct parts of the input.
    - Specify the steps required to complete a task.
    - Provide examples (*few-shot learning* as opposed to *zero-shot* use—this is a form of *in-context learning*).
    - Specify the desired length of the output.

Source: OpenAI, "GPT Best Practices," https://platform.openai.com/docs/guides/gpt-best-practices/six-strategies-for-getting-better-results.

Engineering Tools (2/3)

- *Fine-tuning* of a *foundation* (aka *base*) *model* – Engineer
  - Fine-tuning, aka *domain adaptation*, is the process of tuning a pre-trained model to a specific task.
  - Fine-tuning improves on few-shot learning (by means of the mentioned *prompt design*) by training on many more examples than can fit in the prompt.
  - Once a model has been fine-tuned to the firm's corpus, there is no longer a need to provide examples in the prompt—this saves costs and reduces latency in requests.
  - An insurer may run an entire suite of fine-tuned models, depending on the domain.
    - OpenAI recommends having available "a few hundred high-quality examples" (e.g., reinsurance contracts, claim files, underwriting submissions, etc.) to fine-tune the model to.
  - Later this year, OpenAI will make available for fine-tuning upgraded GPT-3 models as well as GPT-3.5 Turbo, and GPT-4 – current OpenAI base models have been deprecated.[1]

Source: OpenAI, "https://platform.openai.com/docs/guides/fine-tuning," https://platform.openai.com/docs/guides/fine-tuning.
See also Microsoft, "Learn how to customize a model for your application," https://learn.microsoft.com/en-us/azure/cognitive-services/openai/how-to/fine-tuning?pivots=programming-language-studio and AWS, "Fine-tune a foundation model," https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models-fine-tuning.html.
1) The ada, babbage, curie, and davinci models were deprecated on July 6, 2023, and will be turned off effective January 4, 2024.

Engineering Tools (3/3)

- *Embeddings* – Engineer
  - An embedding is a (dense) vector (of numbers) that measures the relatedness of text strings.
    - The distance between two vectors measures their relatedness—small distances suggest high relatedness and large distances suggest low relatedness.
    - Embeddings can be visualized (in boxplots, cluster diagrams, etc.).
  - Embeddings are commonly used for the following tasks (among others) of *text representation*:
    - Clustering (where text strings are grouped by similarity).
    - Anomaly detection (where outliers with little relatedness are identified).
    - Diversity measurement (where similarity distributions are analyzed).
    - Classification (where text strings are classified by their most similar label).
  - There are dedicated embedding models, such as OpenAI's Ada v2.
    - An insurer may run an entire suite of vector spaces of embeddings, depending on the domain.

Source: OpenAI, "Embeddings," https://platform.openai.com/docs/guides/embeddings/what-are-embeddings.

Multimodality

- LLMs are not limited to textual data alone, as they can be pre-trained on…
  - code from various programming languages,
  - mathematical equations,
  - images (and their captions), and
  - audio recordings (and their transcripts).
- Multimodality mimics human cognition and prepares LLMs for the embodiment in robots.
  - As the densest representation of knowledge, text remains central to multimodality.
- Google's PaLM-E is an example of a generalist robotics model based on a multimodal LLM.
  - The inputs to PaLM-E are text and other modalities – images, robot states, scene embeddings, etc.
  - The output is text, which could be an answer to a question, or a sequence of decisions in text form.

See Google Research, *PaLM-E: An embodied multimodal language model*, https://ai.googleblog.com/2023/03/palm-e-embodied-multimodal-language.html.
PaLM-E stands for Pathways Language Model-Embodiment. PaLM is Google's flagship LLM.
PaLM-E is a joint research project with Technical University Berlin.

The Corporate Value Matrix

Unleashing novel data at scale, to be processed in novel workflows

	Legacy Workflow	Novel Workflow
Novel Data	Analytics for Supervised Tasks (e.g., Regression Analysis, Machine Learning, Deep Learning)	Generative AI (e.g., Few-shot learning using prompt design, Fine-tuning of foundation models, Embeddings for text representation; GitHub Copilot, MS 365 Copilot, MS Dynamics Copilot)
Legacy Data	Core IT	Feature Engineering and Machine Learning (e.g., Structured and Unstructured Data Intake, Robotic Process Automation)

Matrix inspired by remarks delivered by Lewis Z. Liu on July 12, 2023, "Digging Deeper: Eigen, IDP, GPT, and all things AI," Eigen Technologies, https://eigentech.com. IDP: Intelligent Document Processing.
For the use of Gen AI in unleashing novel data, see Melissa Dell, *Unleashing Novel Data at Scale*, Spring 2023, https://dell-research-harvard.github.io/blog.html.

Aspects of Debate

Data Privacy Aspects of Popular LLM Tools

Conversational AI Tool	Is Data Used for RLHF?	Chat History Stored in User Account
Google Translate (Free version)	No	No
DeepL (Free version)	Yes	No
DeepL Pro (Paid Version)	No	No risk of data loss if subscription is on an enterprise license.
ChatGPT (Free version, https://openai.com)	Yes	Yes, in a personal OpenAI account.
ChatGPT Plus (Paid version, https://openai.com)	No	Yes, in a personal OpenAI account.
Google Bard	Yes	Yes, in a personal Google account.
Bing Chat / Bing Chat Enterprise (Generative Search)	Yes / No	Search history is stored in the browser. Chat history is not stored.
Google Generative Experience (SGE) (Generative Search)	Yes	Search history is stored in the browser. Chat history is not stored.
Gen AI Service provided by Hyperscaler (e.g., MS Azure OpenAI Service)	No	No risk of data loss as subscription is on an enterprise license.

RLHF: Reinforcement learning by human feedback. RLHF is a technique by which a reward model is trained using human feedback for the purpose of predicting if a given output of the AI system is good (high reward) or bad (low reward). RLHF does not make contributions to the LLM's corpus.

Which Aspects of the Debate are Specific to AI?

Aspect of Debate	Artificial Intelligence	Human Intelligence
Human Agency	Decision-making is surrendered to the machine with unknown consequences.	Where AI is the copilot (rather than the pilot), human agency is preserved.
Explainability	The workings of deep learning models are not fully understood.	The workings of the human brain are not fully understood.[1]
Data Loss Prevention	OpenAI uses trainers to improve the model. There are data harvesting websites, which act as interface between the user and a chat AI engine.	Data harvesting websites that offer pdf conversion or translation services have existed for many years. IT Security has data loss prevention measures are in place.
Hallucination	LLMs tend to hallucinate as the most probable next token is not necessarily the one that maximizes the truthfulness of the generated text. Mitigating factors are prompt design (e.g., chain-of-thought prompting)[2] as well as the RALM architecture[3] applied in generative search where an LLM is augmented by a *retriever model.*	The human brain at times makes erroneous connections, creating false statements, sometimes wittingly.
Fairness	It is now recognized that different concepts of fairness exist and that they are inherently incompatible with one another.[4]	The field of behavioral economics has established that human decision-making is prone to biases, erring systematically in certain contexts.[5]
Disparate Impact	Transparency on disparate impacts, that is, on the error rates of algorithms by groups—group membership is defined by sensitive attributes.	The debate has shifted from fairness to disparate impacts—create transparency on disparate impacts on groups of consumers.
Randomness	An AI system, as machine learning, makes use of random processes and hence does not necessarily deliver the same completion for repeated identical prompts.	Although there is randomness to human thought, humans can create processes that deliver replicable results.
Copyright	LLMs process publicly available (and, potentially, licensed) information, some of which is subject to copyright.	Conversational AI pointing to copyrighted material does not differ from Google search or Bing. Content created using copyrighted material is part of the debate around generative AI in the United States and in the European Union.

1) See for instance the comments by Google CEO Sundar Pichai in CBS News, *60 Minutes, 4/16/2023: Revolution; The Unlikely Adventures of David Grann*, https://www.youtube.com/watch?v=TUCnsS72Q9s.
2) GPT-3.5 Turbo (and other, modern LLMs) have automated chain-of-though prompting by breaking down the user's prompt into a chain of thought in an initial step.
3) RALM stands for retriever-augmented language model. In this architecture, the LLM interacts with an external, retrievable database.
4) See for instance, "Understanding the Importance of Algorithmic Fairness," Gen Re, February 17, 2022, https://www.genre.com/knowledge/blog/2022/february/understanding-the-importance-of-algorithmic-fairness-en.
5) See Kahneman, Daniel, and Amos Tversky (1984) "Choices, values, and frames." *American Psychologist* 39(4), 341–350, https://doi.org/10.1037/0003-066X.39.4.341. Kahneman was awarded the 2022 Nobel Prize in economics. (An early death denied Amos Tversky sharing the prize.)

United States Response

- The U.S. national strategy on AI is currently defined through Executive Orders.[1]
- In October 2022, the Biden administration published a "Blueprint for an AI Bill of Rights," which is generally voluntary but encourages agencies to move AI principles into practice.[2]
- On March 11, 2023, "the Commerce Department put out a public request for comment on… whether potentially risky new AI models should go through a certification process…"[3]
- On June 21, 2023, Senate Majority Leader Chuck Schumer launches the SAFE Innovation Framework, which aims to develop the Senate's policy response and "write the rules of the road on AI."[4]
- Various states, such as California, Connecticut, Illinois, and Texas, have started "to take action to protect the public from the potential harms of these technologies."[5]
- It has been argued that the states' focus is not on AI *per se* but more generally on rules that power automated decision-making of high impact on consumers.[6]

1) See Executive Order 13859 of February 11, 2019, "Maintaining American Leadership in Artificial Intelligence," https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence and Executive Order 13960 of December 3, 2020, "Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government," https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.
2) See https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.
3) See https://ntia.gov/issues/artificial-intelligence/request-for-comments.
4) See https://www.democrats.senate.gov/imo/media/doc/schumer_ai_framework.pdf. See also https://www.csis.org/events/sen-chuck-schumer-launches-safe-innovation-ai-age-csis.
5) Brookings Institution (March 22, 2023) "How California and other states are tackling AI legislation," https://www.brookings.edu/blog/techtank/2023/03/22/how-california-and-other-states-are-tackling-ai-legislation/.
6) *Ibid.*

International Responses

- The Italian Data Protection Authority (*Garante*) initially banned ChatGPT, "saying that OpenAI had 'no legal basis' for using the data it had amassed about Italian residents to train its algorithms and that it was too easy for children to access."[1]
- The EU draft law on AI (passed on June 14, 2023) includes a provision according to which developers of generative AI models will have to publish a 'sufficiently detailed summary of the use of training data protected under copyright law."[2]
- The UK issued a white paper for regulating AI, establishing five principles for "putting the UK on course to be the best place in the world to build, test and use AI technology."[3]
- China's draft regulation requires that AI service providers use data from legitimate sources, effectively control the quality of training data, respect IP rights, and comply with PIPL."[4,5]

1) "ChatGPT Ban Lifted in Italy After Data-Privacy Concessions," *The Wall Street Journal*, April 28, 2023, https://www.wsj.com/articles/chatgpt-ban-lifted-in-italy-after-data-privacy-concessions-d03d53e7.
2) European Parliament (May 11, 2023), AI Act: Compromise Text, https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf.
3) UK Department for Science, Innovation & Technology, *AI regulation: a pro-innovation approach*, March 29, 2023, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach. The consultation on the white paper closes on June 21, 2023. The word "trust" appears 61 times in the document.
4) Cyberspace Administration of China (CAC) (July 13, 2023) *Interim Measures for the Management of Generative Artificial Intelligence Services, http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.*
5) PIPL: Personal Information Protection Law, adopted on Aug. 20, 2021, at the 30th Session of the Standing Committee of the 13th National People's Congress.

Changing Skills Demand

Why Are There Still So Many Jobs?

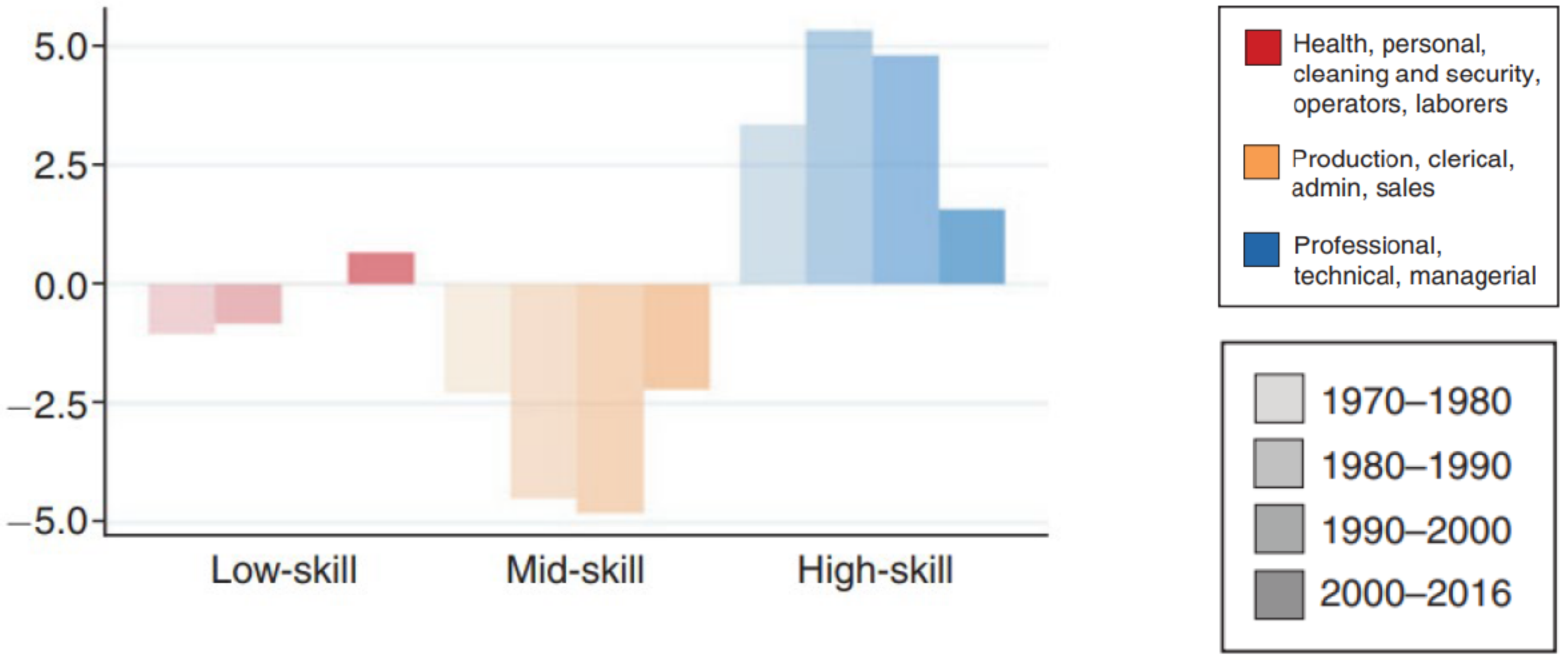
Workplace automation has been going on for two centuries

- Artificial intelligence has led to a resurgence of automation anxiety.
- Automation *substitutes* for labor in performing routine, codifiable tasks.
- Automation also *complements* labor, amplifying the comparative advantage of workers in supplying problem-solving skills, adaptability, and creativity.
- Change in technology changes the types of jobs available, the types of skills in demand to fill these jobs, and what these jobs pay.
- In the past couple of decades, there has been a noticeable polarization of the labor market, in which wage gains went disproportionately to those at the top and at the bottom of the income and skill distribution.

Source: David Autor (2015) "Why Are There Still So Many Jobs: The History and Future of Workplace Automation," *Journal of Economic Perspectives* 29(3): 3-30, https://www.aeaweb.org/articles?id=10.1257/jep.29.3.3.

Change in Demand of Workplace Skills (USA)

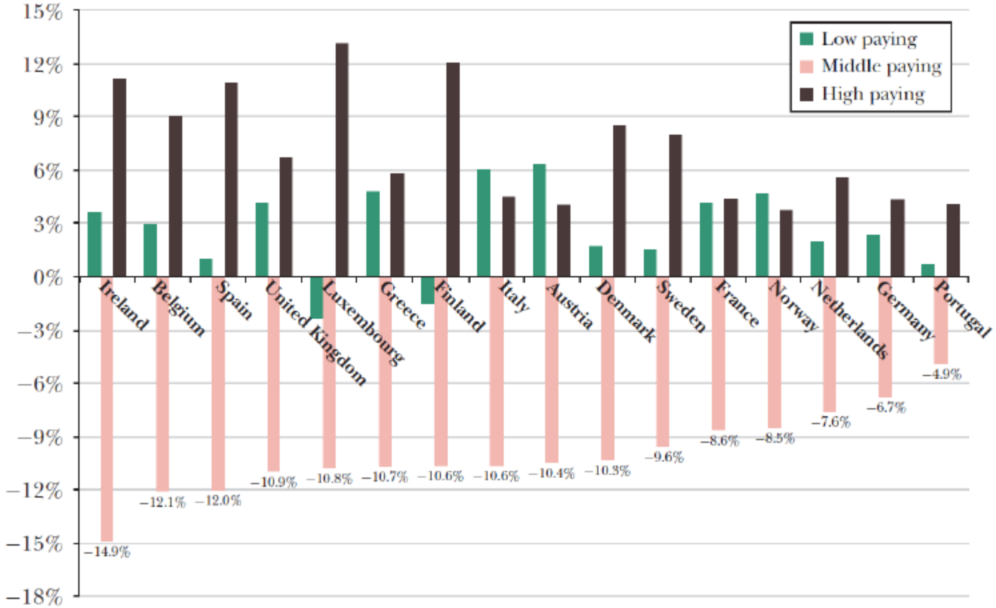
Changes in occupational employment shares among work-age adults, 1980-2016



Source: David Autor (2019) "Work of the Past, Work of the Future," *AEA Papers and Proceedings* 109: 1-32.

Change in Demand of Workplace Skills (Europe)

Change in occupational employment shares in low, middle, and high-wage occupations



16 European Union (EU) Countries, 1993-2010.

Roughly, high-, middle-, and low-paying occupations map into abstract-, routine-, and manual-intensive occupations, respectively.

Notes: High-paying occupations are corporate managers; physical, mathematical, and engineering professionals; life science and health professionals; other professionals; managers of small enterprises; physical, mathematical, and engineering associate professionals; other associate professionals; life science and health associate professionals. Middle-paying occupations are stationary plant and related operators; metal, machinery, and related trade work; drivers and mobile plant operators; office clerks; precision, handicraft, craft printing, and related trade workers; extraction and building trades workers; customer service clerks; machine operators and assemblers; and other craft and related trade workers. Low-paying occupations are laborers in mining, construction, manufacturing, and transport; personal and protective service workers; models, salespersons, and demonstrators; and sales and service elementary occupations.

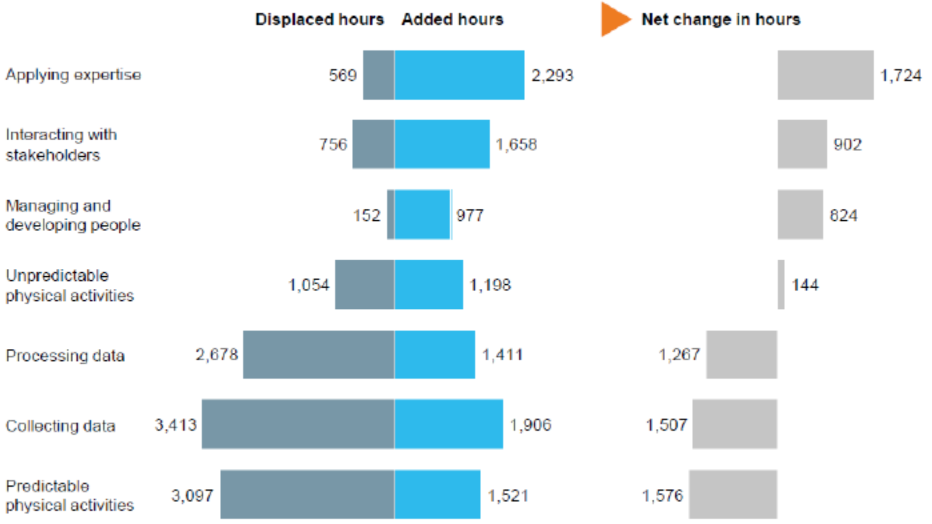
Source: David Autor (2015) "Why Are There Still So Many Jobs: The History and Future of Workplace Automation," *Journal of Economic Perspectives* 29(3): 3-30.

Activities Projected to be Transformed by Automation

Prior to the arrival of generative AI – Germany as an example

Activities within all occupations will shift: New work will involve more application of expertise, interaction, and management

Total hours by activity type, Germany example, 2016–30 (midpoint automation, step-up demand)
Million FTE hours



NOTE: Some occupational data projected into 2016 baseline from latest available 2014 data.

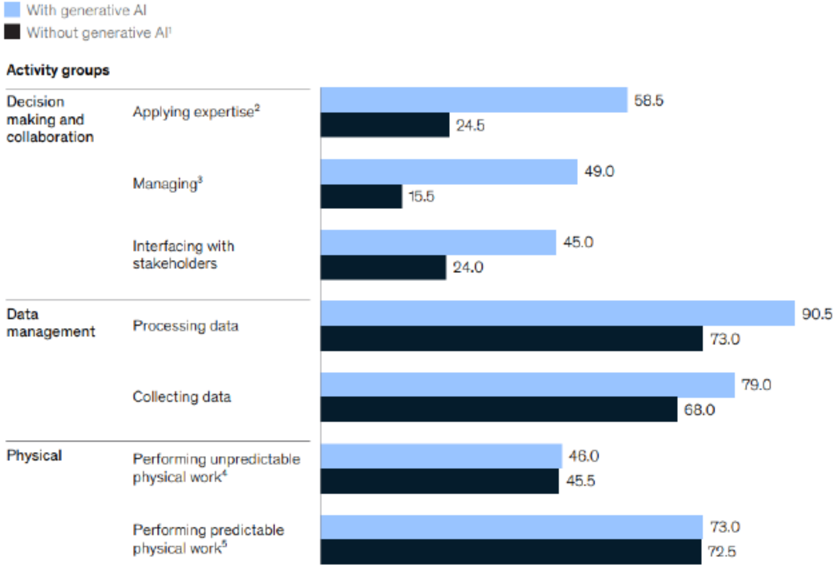
SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

Source: McKinsey & Company (December 6, 2017) "Jobs lost, jobs gained: Workforce Transitions in a time of Automation," www.mckinsey.com.

Activities Projected to be Transformed by Automation

Following the arrival of generative AI

Overall technical automation potential, comparison in midpoint scenarios, % in 2023



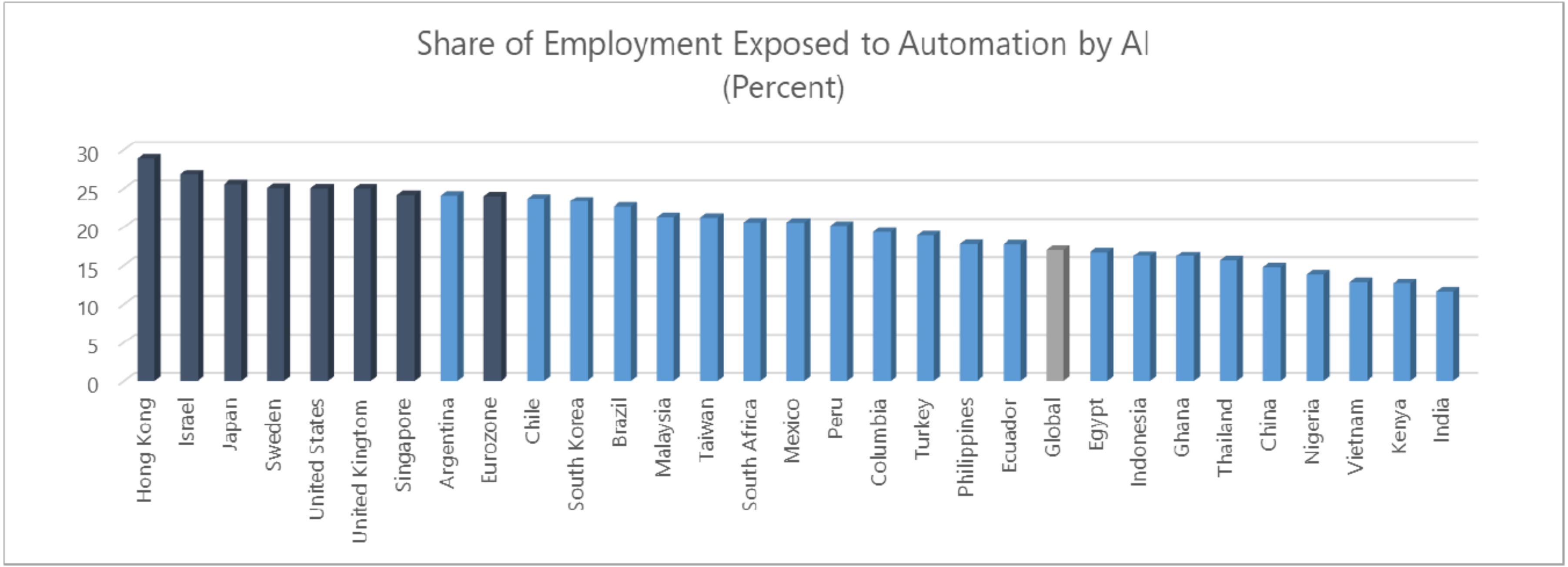
Note: Figures may not sum, because of rounding.
[1]Previous assessment of work automation before the rise of generative AI.
[2]Applying expertise to decision making, planning, and creative tasks.
[3]Managing and developing people.
[4]Performing physical activities and operating machinery in unpredictable environments.
[5]Performing physical activities and operating machinery in predictable environments.
Source: McKinsey Global Institute analysis

- "Without generative AI" projections (dark blue) refer to the 2017 McKinsey study "Jobs lost, jobs gained: Workforce Transitions in a time of Automation," available at mckinsey.com.
- Note that automation not only substitutes for human activities in *decision-making and collaboration* but also complements these activities.

Source: McKinsey & Company (June 2023) "The Economic Potential of Generative AI: The Next Productivity Frontier," www.mckinsey.com.

Employment Exposed to Automation by AI

Projected share of vulnerable jobs is higher in high-income countries

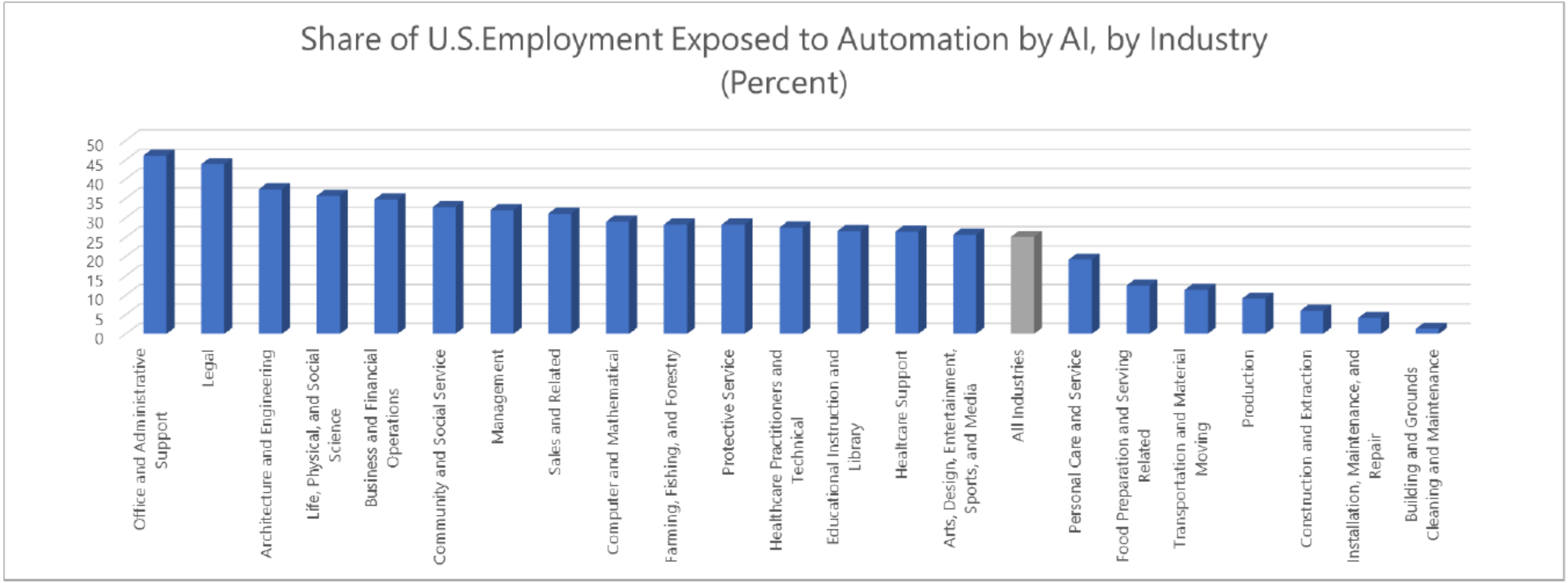


Dark blue: Developed economy. Light blue: Emerging economy.

Data source: Joseph Briggs and Devesh Kodnani (March 26, 2023) "The Potentially Large Effect of Artificial Intelligence on Economic Growth," *Global Economics Analyst*, Goldman Sachs. The study is available from Nasdaq at https://www.nasdaq.com/articles/generative-ai-could-impact-300-mln-jobs%3A-goldman-sachs.

U.S. Employment Exposed to Automation by AI

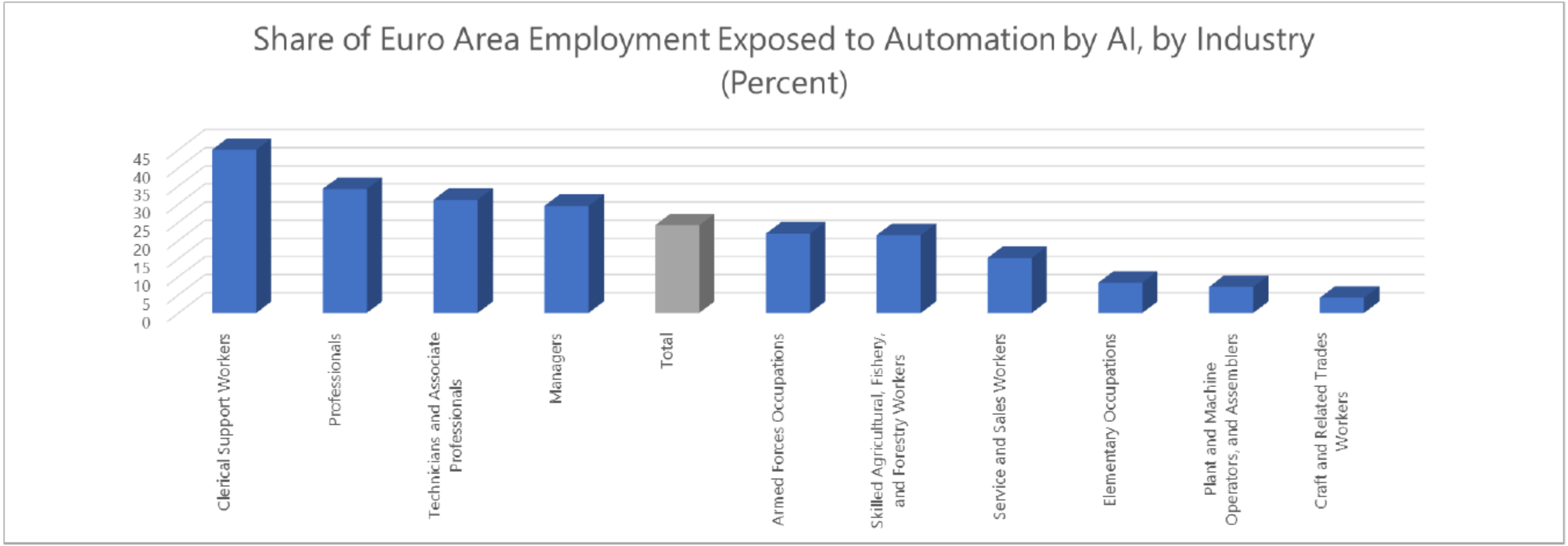
Clerical jobs performing routine, codifiable tasks are most vulnerable



Data source: Joseph Briggs and Devesh Kodnani (March 26, 2023) "The Potentially Large Effect of Artificial Intelligence on Economic Growth," *Global Economics Analyst*, Goldman Sachs. The study is available courtesy of Nasdaq at https://www.nasdaq.com/articles/generative-ai-could-impact-300-mln-jobs%3A-goldman-sachs.

Euro Area Employment Exposed to Automation by AI

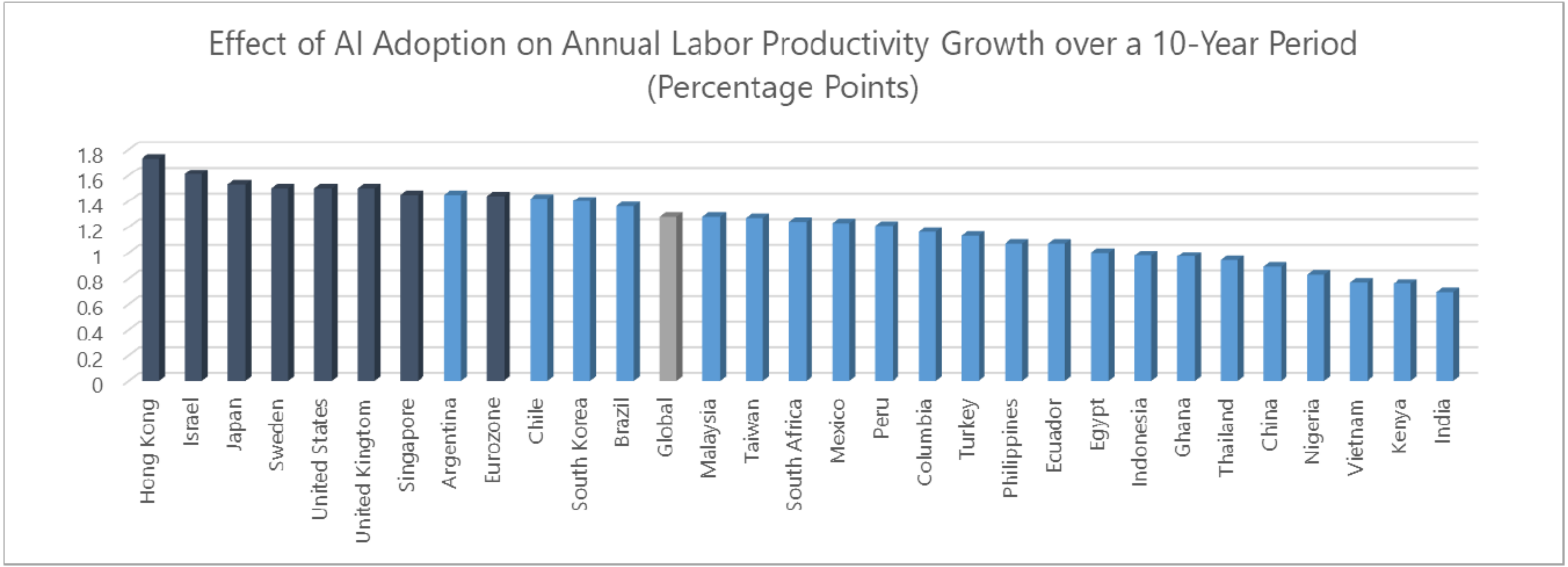
Clerical jobs performing routine, codifiable tasks are most vulnerable



Data source: Joseph Briggs and Devesh Kodnani (March 26, 2023) "The Potentially Large Effect of Artificial Intelligence on Economic Growth," *Global Economics Analyst*, Goldman Sachs. The study is available courtesy of Nasdaq at https://www.nasdaq.com/articles/generative-ai-could-impact-300-mln-jobs%3A-goldman-sachs.

Projected Effect of AI Adoption on Productivity Growth

Potential for a productivity boom – Precedents are the motor car and the personal computer



Dark blue: Developed economy. Light blue: Emerging economy. Productivity booms of modern times are associated with the invention of the motor (ca. 1890; productivity boom in 1917-1932) and the personal computer (1981; 1995-2015). See Martin Wolf (May 9, 2023) "The threat and promise of artificial intelligence," *Financial Times*, www.ft.com.

Data source: Joseph Briggs and Devesh Kodnani (March 26, 2023) "The Potentially Large Effect of Artificial Intelligence on Economic Growth," *Global Economics Analyst*, Goldman Sachs. The study is available courtesy of Nasdaq at https://www.nasdaq.com/articles/generative-ai-could-impact-300-mln-jobs%3A-goldman-sachs.

Ethical Considerations

The Eliza Effect

Source: Rachel Gordon (March 24, 2022) "ELIZA wins Peabody Award," Massachusetts Institute of Technology, https://www.csail.mit.edu/news/eliza-wins-peabody-award.

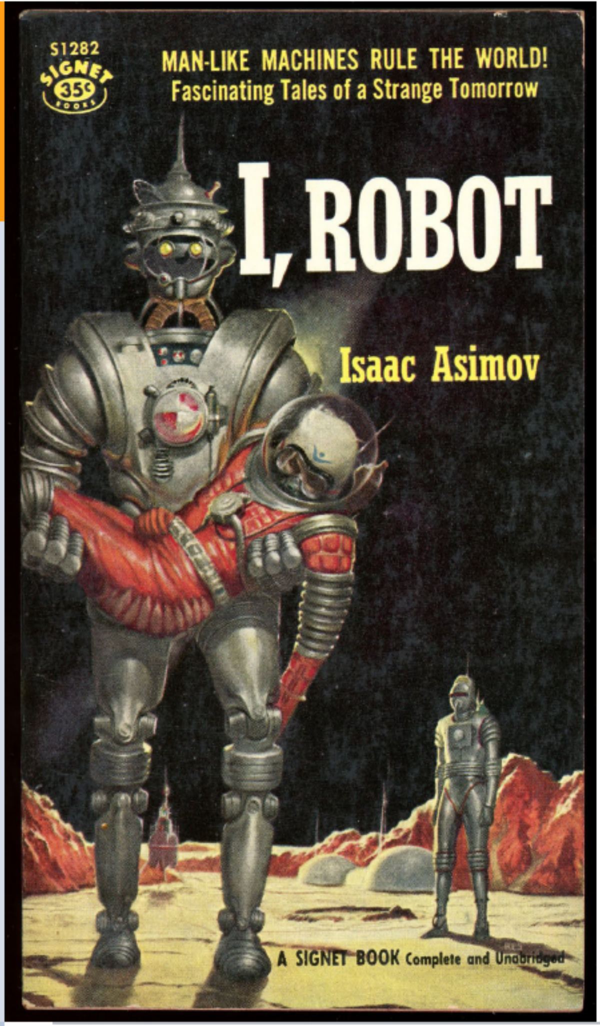
- The Eliza effect refers to people's tendency to attribute human thought processes and emotions to an AI system.
- The Eliza effect was first observed in experimental therapy sessions with an eponymous chatbot created in 1966 at MIT by Joseph Weizenbaum.
- ELIZA interacted with users in typed conversations. It operated by recognizing keywords in a user's statement and then reflecting them back in the form of simple phrases or questions, reminiscent of a conversation with a Rogerian therapist.
- Weizenbaum observed that users got emotionally involved with ELIZA and interacted with it as if it were a human—this phenomenon is known as anthropomorphizing.

Source:
Ellen Glover (March 14, 2023) "What is the Eliza Effect?" Built In, https://builtin.com/artificial-intelligence/eliza-effect. See also Wikipedia, https://en.wikipedia.org/wiki/ELIZA.

The Eliza Effect Revisited

- The anthropomorphizing of computer programs can contribute to social isolation but can also serve as a mitigant to the effects of social isolation.
  - Both aspects are central to the 2013 movie *Her*, where a man who develops a relationship with an artificially intelligent virtual assistant personified through a female voice.[1]
  - Early chatbots that mimic conversations with humans are Replika, Character.AI, and My AI.[2]
  - SAR (Socially Assistive Robotics) is a research area that studies the use of AI for therapeutic purposes (cognitive impairment in older adults, post-traumatic stress disorder, etc.)[3]
- An AI system can impersonate a human without having to pass the Turing test.
  - The ELIZA effect may be exploited as a means of (individual or mass) manipulation.
    - An example of individual manipulation is cyber-attacks via credential theft using social engineering.

1) See Wikipedia, *Her* (film), https://en.wikipedia.org/wiki/Her_(film).
2) "The Other A.I.: Artificial Intimacy With Your Chatbot Friend," *The Wall Street Journal*, August 6, 2023, https://www.wsj.com/articles/when-you-and-ai-become-bffs-ecbcda1e.
3) See for instance the SAR Connect project at Vanderbilt University, https://lab.vanderbilt.edu/rasl/research/sar-connect/.



Asimov's Three Laws of Robotics

Runaround (1942)

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Runaround expounds on challenges that can arise from lexical ambiguity, demonstrates the importance of *prompt design*, and points to the possibility of *hallucinations* ("robotic equivalent of drunkenness").

The short story *Runaround* was first published in 1942 and is included in the collection *I, Robot*, published in 1950, as well as later Isaac Asimov volumes. See https://en.wikipedia.org/wiki/Runaround_(story). The short story is available (among other locations) at https://web.williams.edu/Mathematics/sjmiller/public_html/105Sp10/handouts/Runaround.html.

Keep up with the latest industry trends...

Visit genre.com/perspective

Follow us on:

linkedin.com/company/gen-re

twitter.com/Gen_Re