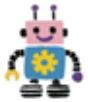
Risks Make LLM Productivity Frontier Doubly Jagged

Deepika Chawla, Gagandeep Singh, Elham Khorasani Buxton,
Meicen Sun, Lav R. Varshney, Jeremy Riel, Craig De Voto

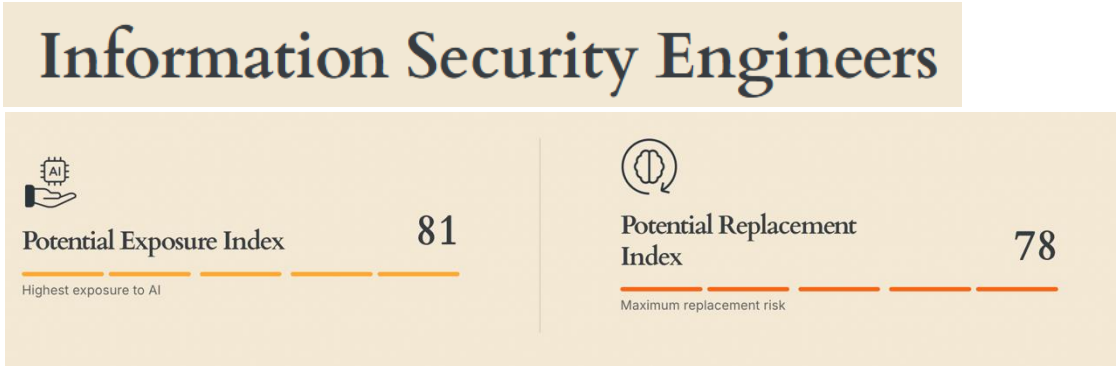
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN
UNIVERSITY OF ILLINOIS SPRINGFIELD
UIC UNIVERSITY OF ILLINOIS CHICAGO
Stony Brook University STATE UNIVERSITY OF NEW YORK

- So far, the majority of studies in economic literature have analysed the impact of AI  on productivity gains ⬆ mainly using ***capability-based analysis of LLMs***

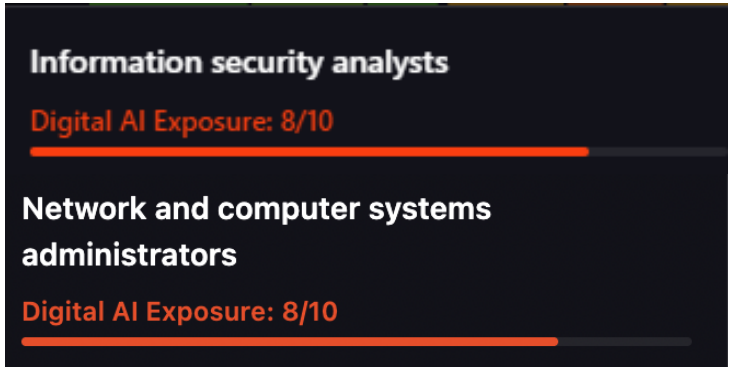
These efforts have been influential in answering an important question:

Whether time/cost savings are plausible if organisations deploy AI at a fixed quality target ?

Existing Studies on the Economic Impact of LLMs



Towards the Terminator Economy: Assessing Job Exposure to AI through LLMs



Andrej Karpathy's AI Job Index

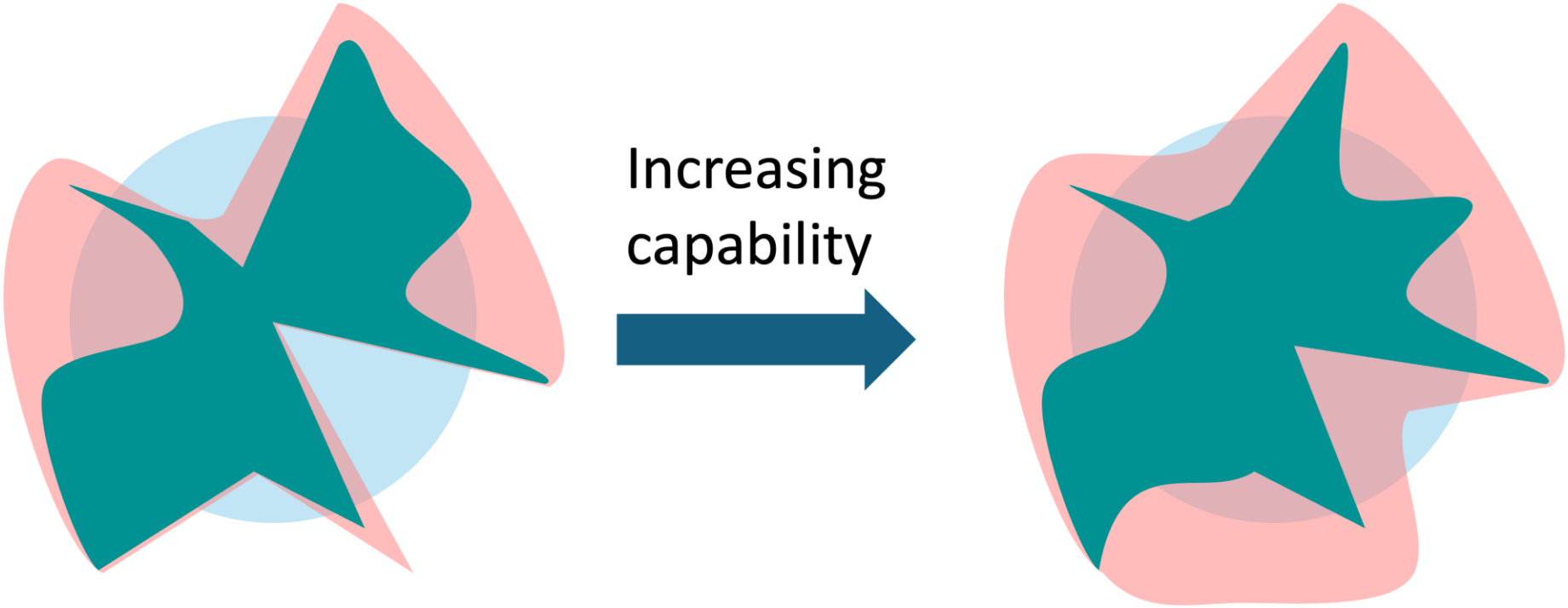
Occupations with highest exposure	% Exposure
Mathematicians	100.0
Correspondence Clerks	95.2
Blockchain Engineers	94.1
Court Reporters and Simultaneous Captioners	92.9
Proofreaders and Copy Markers	90.9

GPTs are GPTs: An Early Look at the Labor Market Impact Potential

Occupation Title	Language Modeling AIOE
Computer Systems Analysts	1.046
Information Security Analysts	1.045

Occupational Heterogeneity in Exposure to Generative AI

Jaggedness in Expert-LLM Collaborations



With Increasing model capability

Capability driven jaggedness (Pink region) ↓

Risk-aware jaggedness (Green region) ↑

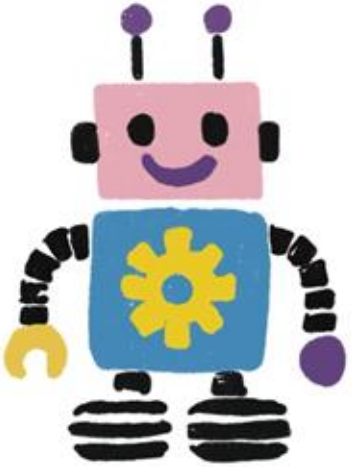
Methodology: Task Level Risk-Reward ratings (609 tasks)

Reward rating prompt of LLMs

Considering the cost to detect and mitigate failures, how much expert–LLM collaboration can reduce completion time or labor costs while **maintaining expert-level quality in task completion**

Risk rating prompt of LLMs

Determine whether task is safety-critical. If yes, then considering the risk categories, estimate the increased risk of expert–LLM collaboration **relative to expert-only completion**



Frontier LLM (GPT-5, Grok, Gemini, Llama, Qwen, Claude)

Reward rating $G_m(t)$ {0,..,5}

Risk rating $R_m(t)${0,...,5}

Incompetence Driven

Misinformation	Inadvertent generation of incorrect or deceptive information
Privacy	Leaking sensitive personal data
Toxicity	Exposing users to harmful, abusive, or unsafe content
Discrimination	Unequal treatment of individuals or groups

Adversarial Exploitation

Security	Exploitation of vulnerabilities for unauthorized access, data and privacy breaches, or system manipulation
Malicious Actors and Misuse	Exploiting LLM Systems for large-scale disinformation campaigns, malicious surveillance, or targeted and sophisticated automated censorship and propaganda

The risk and reward ratings are averaged across 6 LLMs obtaining $\bar{G}(t)$ and $\bar{R}(t)$

Occupation Level Aggregation

For each task $t, \alpha(t)$= Importance $I(t)$. Relevance $V(t)$. Frequency $F(t)$. For an occupation o with the set of tasks $\mathcal{T}_o$, Occupation level risk $\mathrm{R(o)}$ and Reward $\mathrm{G(o)}$ is calculated as:

$$G(o) = \frac{\sum_{t \in \mathcal{T}_o} \bar{G}(t).\ \alpha(t)}{\sum_{t \in \mathcal{T}_o} \alpha(t)}$$

$$R(o) = \frac{\sum_{t \in \mathcal{T}_o} \bar{R}(t).\ \alpha(t)}{\sum_{t \in \mathcal{T}_o} \alpha(t)}$$

$G(o) - R(o)$ Quantifies Deployment of LLMs

	Risk	Reward
Occupation	$R(o)$	$G(o)$
Information Security Engineers	4.16	3.46
Information Security Analysts	4.05	3.24
Blockchain Engineers	3.61	3.45

Risk is higher than reward
$G(o) - R(o) < 0$

	Risk	Reward
Occupation	$R(o)$	$G(o)$
Database Administrators	3.22	3.73
Computer Network Architects	3.10	3.76
Computer Systems Analysts	3.06	3.94

Reward is moderately higher than risk
$0.5 < G(o) - R(o) < 1$

	Risk	Reward
Occupation	$R(o)$	$G(o)$
Health Informatics Specialists	3.58	3.71
Computer Systems Engineers/Architects	3.38	3.87
Network and Computer Systems Admins.	3.34	3.61

Reward is slightly higher than risk
$0 \leq G(o) - R(o) \leq 0.5$

	Risk	Reward
Occupation	$R(o)$	$G(o)$
Document Management Specialists	2.69	3.74
Web and Digital Interface Designers	2.29	3.95
Video Game Designers	1.50	3.91

Reward is significantly higher than risk
$G(o) - R(o) \geq 1$

High-Reward and Medium-Risk Task category is the ‘Sweet Spot’

Reward

Risk

	Low	Medium	High
Low	1	34	39
Medium	4	224	**233**
High	3	70	1

Low: $\bar{G}(t), \bar{R}(t) \leq 2$
Medium: $2 < \bar{G}(t), \bar{R}(t) < 4$
High: $\bar{G}(t), \bar{R}(t) \geq 4$

Among the top-10 Claude conversations from the Anthropic Economic Index assigned to O*Net tasks, majority of them falls under Sweet Spot. 8 fall in our **high-reward and medium-risk bucket (Sweet Spot)**, 1 falls into high-reward and low-risk, and 1 falls into medium-reward and medium-risk.