



Working Paper (WP/25-12)

Labor Market Exposure to AI: From GenAI to Future AGI

Xianguo Huang

November 2025

Disclaimer: The findings, interpretations, and conclusions expressed in this material represent the views of the author(s) and are not necessarily those of the ASEAN+3 Macroeconomic Research Office (AMRO) or its member authorities. Neither AMRO nor its member authorities shall be held responsible for any consequence from the use of the information contained therein.

[This page is intentionally left blank]

Labor Market Exposure to AI: From GenAI to Future AGI

Prepared by Xianguo (Jerry) Huang^{*†}

Authorized for distribution by Abdurrohman (Deputy Director)

November 2025

Abstract

This paper adopts and extends a task-based framework leveraging Large Language Models (LLM) to assess occupational exposure to Generative Artificial Intelligence (GenAI). It reveals cross-country variation by income level, with higher-income economies experiencing greater task-level exposure. At present, the potential for GenAI to augment human work outweighs its automation risks, though this balance may shift with the advent of more advanced AI systems and their integration with other emerging technologies. The paper further conducts an exercise to assess occupational exposure under a scenario where artificial general intelligence (AGI) is realized. Using a combination of LLM-based task evaluations and detailed labor force employment data, the study also presents a country case study. While overall exposure remains moderate, occupations in clerical, administrative, and financial services are particularly vulnerable to GenAI-driven transformation in Brunei. Exposure is notably higher, among individuals with mid-level educational attainment, as well as among women.

JEL codes: J23, J24, J68, O30, O33

Keywords: GenAI, AGI, Labor Market, ISCO-08, Brunei

^{*}Email address: jerry.huang@amro-asia.org.

[†]I thank Laura Britt Fermo, Allen Ng, Toan Long Quach, and Yunhui Zhao for their constructive feedback and suggestions. All remaining errors or omissions are my own responsibility.

Contents

1	Introduction	1
2	GenAI Automation Exposure- Task and Occupation	5
2.1	Task Description Text Data	5
2.2	Methodology	5
2.3	Result for GenAI Exposure	7
2.3.1	Task-level Score: TE-GenAI	7
2.3.2	Occupation-level Score and Classification: OE-GenAI	8
2.4	Beyond GenAI	13
3	A Country Case	14
3.1	OE-GenAI in Brunei	14
4	Discussions	17
4.1	Implications	17
4.2	Suggestions for labor statistics	18
4.3	Limitations and Areas for Future Research	19
5	Conclusions	20
A	Appendix	21
A.1	Models and Prompts	21
A.2	Additional Information of TE-GenAI	23
A.3	Additional Information of OE-GenAI	28
A.4	A Experiment of Country Specification in Prompt	29
A.5	Understanding of GPT, GenAI and AGI by Grok	31
A.6	Examples of AI-related Survey Questionnaire	32

List of Figures

1	AI Diagram	4
2	Statistics of Task-level Exposure Score	9
3	OE-GenAI Score and Classification	11
4	OE-AGI Occupation Risks	13
5	OE-GenAI Classification, by Economic Sector	15
6	OE-GenAI Classification, by Education and Gender	16
A1	Statistics of TE-GenAI, Time Allocation	23
A2	Statistics of OE-GenAI: Density	28
A3	OE-GenAI: "Brunei, a high-income country" vs "a high-income country"	29
A4	OE-GenAI Classification, based on "Brunei, a high-income country"	30

List of Tables

1	Number of ISCO-8 4-digit Occupations and Tasks	5
2	Occupation Classification Criteria Based on TE-GenAI Scores	7

3	Paired t-Test for Task Exposure Score	9
4	Occupation Exposure to GenAI with Automation Potential	12
A1	System and User Prompts	21
A2	Paired T-Test for Task Time Allocation	23
A3	Occupation Task Exposure to GenAI, Highest Group (score=0.95)	24
A4	Occupation Task Exposure to GenAI, Lowest Group (score=0)	26
A5	Example of AI Exposure Scores and Justifications: Domestic Housekeepers (5152), "High-Income Country" vs "Brunei"	27

1 Introduction

The evolution of technology has consistently shaped labor markets throughout history. Recent advancements in Artificial Intelligence (AI) have not only pushed technological boundaries but are also fundamentally transforming the nature of work, prompting extensive research into its implications. The impact of AI on the labor market is typically analyzed through two key dimensions: job automation and displacement, and productivity enhancement. Earlier studies have provided a wide range of estimates regarding automation's potential scope. For example, [Frey and Osborne \(2017\)](#) categorized occupations by their susceptibility to computerization, concluding that nearly half of U.S. jobs are at risk. Conversely, [Arntz, Gregory, and Zierahn \(2016\)](#) employed a task-based methodology and suggested a milder effect, estimating that approximately 9% of jobs across OECD member countries are automatable. While these analyses differ in scale, there is a broad agreement that workers in routine-intensive roles are particularly vulnerable to displacement, while new job opportunities arise. Nevertheless, the net impact on employment remains uncertain.

On the productivity side, consensus seems to be stronger regarding AI's potential benefits. By automating repetitive tasks and enhancing human capabilities, AI can significantly improve production efficiency. Projections indicate that AI could add 14% to global GDP by 2030—approximately \$15.7 trillion ([PWC 2017](#)). Recent attention has turned more to Generative AI (GenAI), a subset of AI capable of creating novel text, images, music, and other content. Large Language Models (LLMs) such as ChatGPT, Grok, and DeepSeek are at the forefront of this shift, offering growing multimodal sophistication and expanding applications. These technologies are redefining job functions by automating routine tasks and amplifying human creativity. For instance, ChatGPT has demonstrated substantial productivity gains, reducing task completion time by 40% and increasing output quality by 18% ([Noy and Zhang 2023](#)). Similarly, tools like GitHub Copilot have expedited software development processes ([Peng et al. 2023](#)).

Distinct from earlier AI automation, which primarily targeted routine tasks, generative AI encroaches on domains requiring imagination and artistry. For example, GenAI can generate creative content like artwork or writing drafts, challenging roles in creative industries. Tacit knowledge and creative judgment—long thought to be insulated from automation—are now partially replicable through large-scale pattern recognition ([Noy and Zhang 2023](#)). This shift could restructure creative occupations over time, with roles increasingly centered on curating, editing, and guiding AI outputs—often termed “prompt engineering” or AI orchestration—rather than traditional creation from scratch.

Studies on GenAI-driven automation have proliferated. One approach is to follow the traditional AI exposure measure linking AI applications and human abilities ([Edward Felten, Raj, and Seamans 2021](#)), extended to calculations covering language modelling ([Ed Felten, Raj, and Seamans 2023](#)). On the other hand, innovative approaches have also emerged with rising popularity, which involve

the use of Large Language Models (LLMs)—the text-based GenAI models themselves equipped with extensive knowledge and reasoning capabilities—to assess automation risks at both task and occupational levels. The application of LLMs as evaluation tools has been adopted across various fields in recent years, with their effectiveness undergoing rigorous scrutiny. Initiated by [Eloundou et al. \(2023\)](#), this stream has been enriched by [Gmyrek, Berg, and Bescond \(2023\)](#), which focuses on a global perspective, and [Chen et al. \(2023\)](#), exploring the potential impacts on the Chinese labor market. In 2024, [Colombo, Mercorio, and Mezzanzanica \(2024\)](#) employed LLM evaluation methods to analyze AI exposure in the U.S. labor market, offering valuable insights into the economic impact of AI-driven automation. Similarly, [Gmyrek, Lutz, and Newlands \(2024\)](#) compared occupational evaluations conducted by GPT-4 against human assessments in the UK, shedding light on LLMs' reliability in workforce analysis. Separately, there is rich literature studying job impact rather than exposure. For instance, [Acemoglu and Restrepo \(2022\)](#) show that automation technologies expand the role of capital and lead to certain jobs displacement, which results in rising wage inequality.

Using GPT as an evaluation tool has also been applied across fields, and its effectiveness has been examined, such as in responses from AI-powered medical chatbots ([Tan et al. 2024](#)), e-commerce product reviews ([Roumeliotis, Tselikas, and Nasiopoulos 2024](#)), and climate change assessments ([Joe, Koneru, and Kirchhoff 2024](#)), or in the financial and economic areas such as in [Bybee \(2023\)](#), where comparisons are made with results from human experts. As an alternative approach, [Handa et al. \(2025\)](#) analyzed a large set of Claude.ai conversations to gain insights into how GenAI is used for automation and augmentation of tasks.¹

This paper builds on the above approach by leveraging recent advances in GenAI, particularly the emergence of more capable models that can process complex instructions, perform reasoning, and access information with minimal latency. It employs Grok (xAI's real-time model) to construct sets of exposure scores for all ISCO-08 tasks conditional on country income-levels and subsequently evaluate the potential for occupation augmentation and automation at an aggregated occupational level. The paper then reassesses these findings within a specific country context by using labor employment at a disaggregated level (using Brunei as a case study). The paper draws on granular labor statistics segmented by economic sector, education level, and gender, providing deeper insights into prospective impacts on the labor market. Brunei was selected due to its unique profile as a high-income, oil-dependent economy in Southeast Asia, offering insights into how AI might aid economic diversification in resource-reliant settings. Separately, a predictive exercise is conducted with regard to the automation risks of Artificial General Intelligence (AGI), a future hypothetical technology. AGI refers to a hypothetical form of AI that possesses the ability to understand, learn, and apply knowledge across a broad range of tasks, exhibiting more general reasoning, adaptability, and problem-solving, closer to how humans handle diverse challenges (Figure 1).

1. For the use cases of LLM advancements in the field of economics, [Korinek \(2024\)](#) provides a comprehensive coverage.

The remainder of this paper is organized as follows. Section 2 presents the LLM-based results, including task- and occupation-level exposure. Section 3 narrows the focus to a specific country context. Section 4 expands on these findings and discusses their broader implications. Finally, Section 5 concludes the paper.

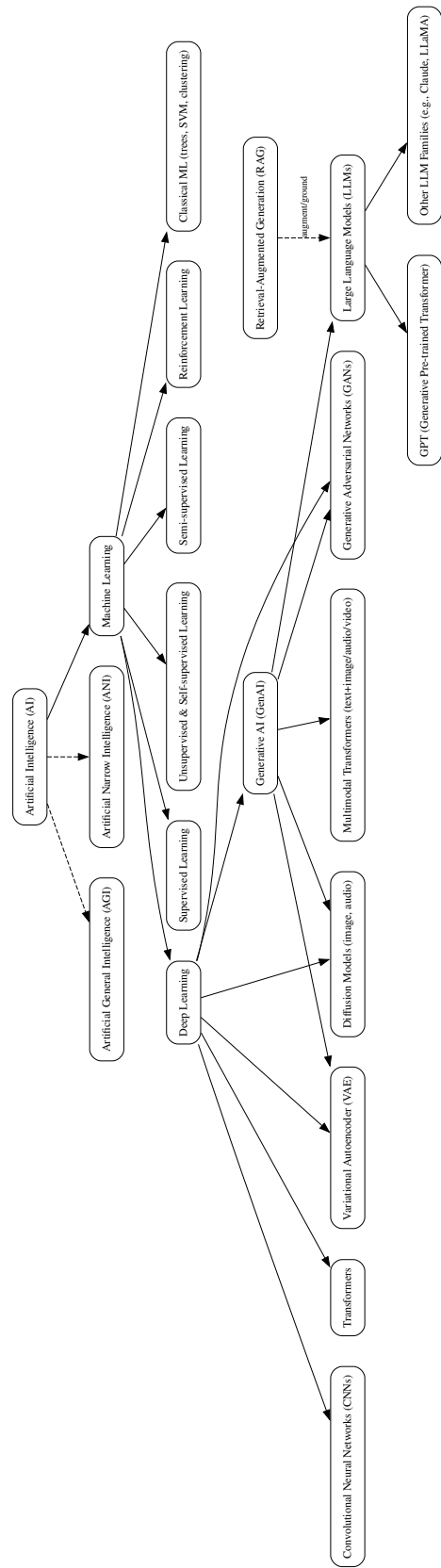


Figure 1: AI Diagram

Source: author's compilation.

2 GenAI Automation Exposure- Task and Occupation

2.1 Task Description Text Data

The ISCO (International Standard Classification of Occupations) data at the 4-digit level encompasses 427 occupations (excluding armed forces occupations) with 3,265 tasks (ILO 2023).² Instead of focusing on LLM-generated tasks to balance the number of tasks across professions (Gmyrek, Berg, and Bescond 2023), this paper focuses on ISCO tasks directly provided by the ILO to minimize potential biases introduced by LLM task generation (Table 1).

Table 1: Number of ISCO-8 4-digit Occupations and Tasks

ISCO 1-digit	Category	Occupation	Task
1	Managers	30	293
2	Professionals	92	773
3	Technicians and associate professionals	82	586
4	Clerical support workers	29	163
5	Service and sales workers	39	273
6	Skilled agricultural, forestry and fishery workers	18	176
7	Craft and related trades workers	65	503
8	Plant and machine operators, and assemblers	39	286
9	Elementary occupations	33	212
Total		427	3265

Source: ILO 2023.

2.2 Methodology

As discussed earlier in Section 1, the use of large language models (LLMs) as expert evaluators has gained significant traction in recent years, driven by their rapidly increasing capabilities. This potentially provides a new research paradigm, including in latest studies leveraging LLM-based assessments to estimate AI exposure across labor markets.

Given the diverse range of LLMs available as potential evaluators, the choice of model necessitates careful consideration of several influential factors. First, each LLM is trained on datasets with distinct scopes, timeframes, and potential regional biases, all of which can influence its performance. While the impact of outdated training data is not explicitly quantified in existing studies, models with earlier knowledge cutoffs may struggle to provide accurate assessments in contexts where GenAI adoption is nascent or has evolving granularity. Conversely, newer models, often trained on datasets

2. The count excludes three professions classified under Code 0 (military) and six profession titles labeled as "Not Elsewhere Classified" across various groups. The resulting total number of ISCO tasks—based on 427 professions—is 3,265, which exceeds the 3,123 ISCO tasks reported in Gmyrek, Berg, and Bescond (2023).

disproportionately focused on technology-intensive economies, may overestimate automation risks when applied to regions with slower AI adoption.

Second, architectural advancements—such as larger parameter counts, enhanced reasoning abilities, and improved contextual comprehension—can enhance an LLM’s ability to analyze complex task descriptions and justify automation scores. Recent innovations in transformer architectures, including fine-tuning techniques like reinforcement learning from human feedback (RLHF), have been shown to improve LLMs’ ability to handle nuanced and multifaceted tasks, thereby increasing assessment accuracy in complex scenarios ([Steyvers et al. 2025](#)).

Third, an effective evaluator must possess a sufficiently large context window to process extensive input instructions via prompts. A larger context window enables the model to jointly analyze multiple tasks within a single request, thereby ensuring a more holistic evaluation of automation potential across occupations. Taking these considerations into account, we selected Grok-2, a model developed by xAI, which incorporates real-time data—a distinct advantage over models with knowledge cutoffs with older dates, and yet with a sufficiently large context window to handle our processing requests. To ensure deterministic outputs, the model’s temperature parameter³ is set to zero, instructing it to select the most probable next token in each response.

Both system and user prompts were provided to ensure more calibrated and structured responses from the chosen LLM. A system prompt guided the model in adopting the role of a skills specialist with expertise in AI technologies. The user prompts instructed the task evaluation process such that the LLM assessed an entire list of tasks associated with each occupation in a single API request to the cloud server. As job descriptions and task statements were processed in batches, it also helped optimize API efficiency and minimize latency. The prompts not only required the generation of automation scores (ranging from 0 to 1, with higher scores indicating greater automation potential) but also elicited the model to estimate the proportion of work time allocated to each task with the consistency cross-checked in the batch processing of tasks.

Meanwhile, leveraging recent advancements in reasoning models and the availability of larger context windows, this paper incorporated additional instructions to promote consistency and enhanced the rigor of the justification process, while using various prompt techniques to minimize hallucination. In particular, it adopted the Chain of Thought (CoT) prompting technique ([Chen et al. 2023](#)) to encourage structured and sequential reasoning. Furthermore, the paper applied the "tip and penalize" principle—an incentive-based strategy that explicitly defines rewards for well-reasoned responses and penalties for inadequate ones—as demonstrated by [Bsharat, Myrzakhan, and Shen \(2023\)](#) and [Chen and Zhao \(2024\)](#). Few-shot prompting⁴ with a small number of examples also

3. The temperature parameter in large language models (LLMs) controls the randomness of the model’s output generation.

4. Few-shot prompting is a technique used in large language models (LLMs) where a small number of examples are included in the input prompt to demonstrate the desired output format or reasoning process.

helps guide its response to queries. Appendix Table A1 provides further details on the prompting structure. Despite these measures, potential hallucinations remain a risk, mitigated here by setting temperature to zero and requiring justifications.

Based on the disaggregated results of Task Exposure to GenAI automation potential (TE-GenAI) and the corresponding time allocation across tasks within each occupation, this paper defines the weighted average of task-level scores as the overall Occupation Exposure to GenAI automation potential (OE-GenAI). Building on the approach of [Carbonero et al. \(2023\)](#) and [Gmyrek, Berg, and Bescond \(2023\)](#), which categorizes exposure levels using simple averaging, this analysis instead applies a weighted mean and standard deviation, aggregating TE-GenAI scores based on task-specific time shares.

The resulting first and second moments of OE-GenAI are then used to classify occupations according to their automation and augmentation potential. While the classification relies on rule-of-thumb threshold values and leads to a large category which is unclassified, such categories offer an intuitive framework for interpreting occupational risks derived from disaggregated task-level exposure scores, as summarized in Table 2. The low-mean groups (<0.4) include those with augmentation potential or not affected, and the high-mean groups (>0.6) are divided into those with automation potential or big unknown. The rest with average means (0.4 to 0.6) are "others" left not classified. In [Gmyrek, Berg, and Bescond \(2023\)](#), three groups "not affected", "big unknown" and "others" are classified as "Other occupations".

Table 2: Occupation Classification Criteria Based on TE-GenAI Scores

Category	Criterion 1	Criterion 2
Augmentation potential	< 0.4	$\mu + \sigma > 0.5$
Automation potential	> 0.6	$\mu - \sigma > 0.5$
Not affected	< 0.4	$\mu + \sigma \leq 0.5$
Big unknown	> 0.6	$\mu - \sigma \leq 0.5$
Others	Other unspecified	

Note: mean μ and $\mu \pm \sigma$ refer to mean and standard deviation.

2.3 Result for GenAI Exposure

2.3.1 Task-level Score: TE-GenAI

Figure 2 presents key statistics and the density function of TE-GenAI, illustrating the extent to which job tasks are exposed to GenAI across countries with low-, middle-, and high-income classifications. The corresponding mean automation scores are 0.453, 0.427, and 0.404 for high-, middle-, and low-income groups, respectively, indicating a positive correlation between higher income levels and increased GenAI exposure. This finding is consistent with [Gmyrek, Berg, and Bescond \(2023\)](#), which suggests the share of employment subject to automation risks is higher in high-income countries.

Notably, the distribution for high-income countries exhibits a greater probability mass toward the right-hand side, suggesting that tasks in these economies are more susceptible to automation. This pattern aligns with the notion that high-income countries tend to have a greater concentration of roles involving cognitive, repetitive, or structured tasks—where GenAI could excel. Additionally, the density function reveals a pronounced peak around the 0.6 score range across all income groups, highlighting a strong tendency for LLM-based scoring to cluster around this threshold. This could reflect an inherent model bias or a structural characteristic of job tasks that commonly receive intermediate exposure ratings.

While the aggregate statistics suggest that the average scores differ across income groups, it is hard to conclude whether there are systematically different. A comparison at a task level across groups is conducted by using a Paired t-Test, which shows such scores do differ from each other in a systematic way, suggesting LLM scoring does take into account the context of income differences (Table 3). The same examination is also made on data of time allocation for each task. However, there are no systematic differences in time allocation driven across different income country groups (Appendix A.2). This may indicate that task time shares are relatively universal across economies, while automation feasibility varies with technological infrastructure and adoption rates.

In terms of scores at the tail sides, the LLM provides relatively consistent results for those tasks highest and least exposed to automation, where the tasks with the highest automation exposure score (0.95) fall into the following categories: i). *operating switchboards and consoles* by telephone switchboard operators in high-income and middle-income countries; ii). *recording notes for follow-up actions, updating marketing databases, and maintaining call statistics* by contact centre salespersons in high-income and middle-income countries; and iii). *issuing tickets, passes and vouchers* by travel consultants and clerks in middle-income countries. In contrast, in lower-income countries, only *maintaining call statistics* by contact centre salespersons receives the highest automation score.

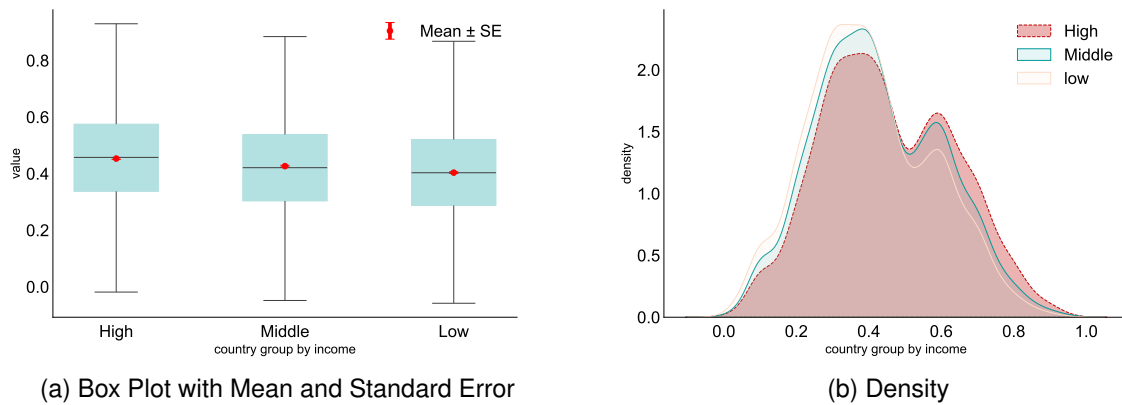
On the other end of the spectrum, tasks with the lowest score of TE-GenAI (0.00) are consistently associated with *athletes and sports players* across all country groups, reflecting the nature of physical, real-time decision-making tasks that remain beyond GenAI capabilities. The reasoning of LLMs for such scoring is fairly balanced, stressing supportive arguments and constraints that GenAI faces. Appendix Table A3 and A4 provide more details for the task groups with highest and lowest scores, including additional examples and LLM justifications to illustrate reasoning depth.

Table 3: Paired t-Test for Task Exposure Score

Pair	T-statistic
High vs. Middle Income	18.14***
High vs. Low Income	30.86***
Middle vs. Low Income	16.0***

Note: ° :> 0.1; *:< 0.1; **: < 0.05; ***:< 0.01.

Figure 2: Statistics of Task-level Exposure Score



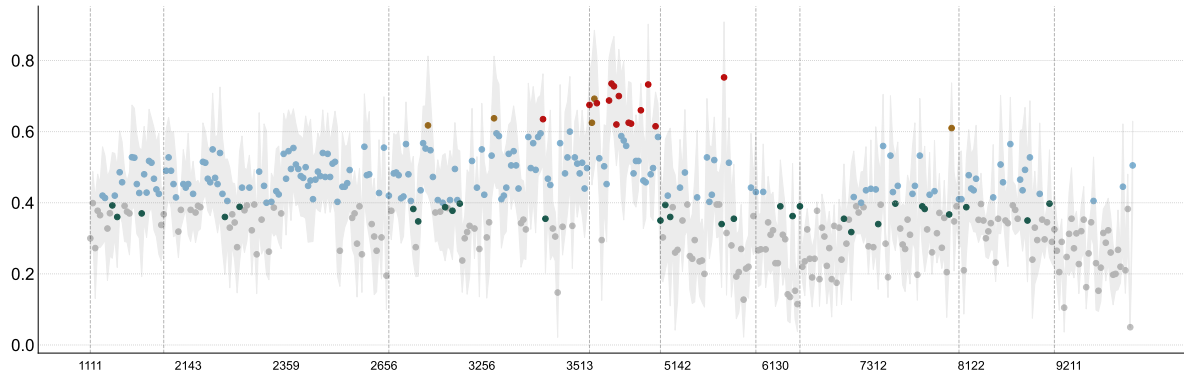
2.3.2 Occupation-level Score and Classification: OE-GenAI

Figure 3 displays the classification of occupations based on the criteria defined in Table 2, distinguishing between augmentation potential, automation potential, and other categories. Overall, the patterns are broadly similar across country groups. Occupations with ISCO-08 codes greater than 5142—typically representing service and sales workers, as well as agricultural and elementary occupations—tend to exhibit lower exposure scores.

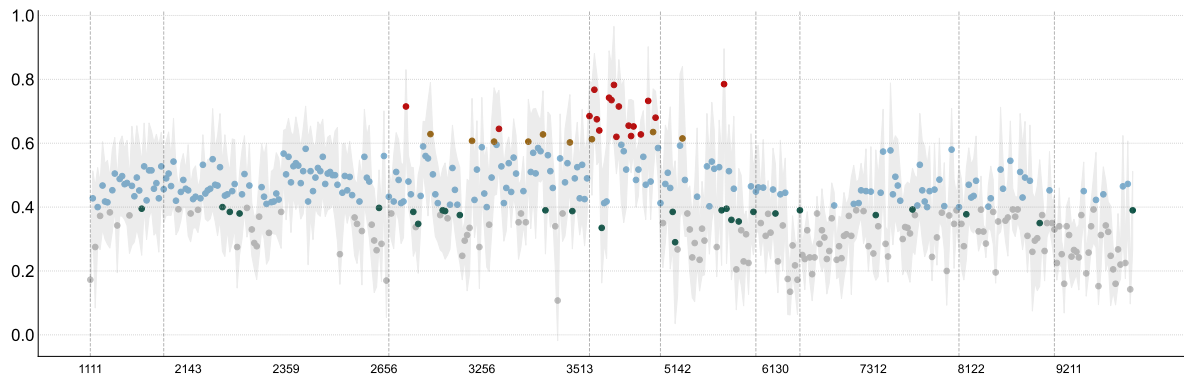
That said, due to systematic differences in task-level exposure scores (TE-GenAI), as discussed in Subsection 2.3.1, the resulting occupation-level exposure (OE-GenAI) shows notable variation across country income groups. Higher-income countries exhibit a greater number of occupations classified under both augmentation and automation potential, suggesting a broader impact of GenAI at the occupational level.

Table 4 provides further detail on occupations subject to automation risk. In high-income countries, 6.1 percent of professions fall under the automation potential category, while 8.0 percent are classified under augmentation potential. These shares are lower in middle-income countries, at 4.2 percent and 6.3 percent, respectively. In low-income countries, the proportions are 3.3 percent and 6.8 percent. Notably, certain occupation groups—such as clerical professions—consistently face higher automation risks across all income groups. Appendix Figure A2 provides density functions for occupation scores for further information.

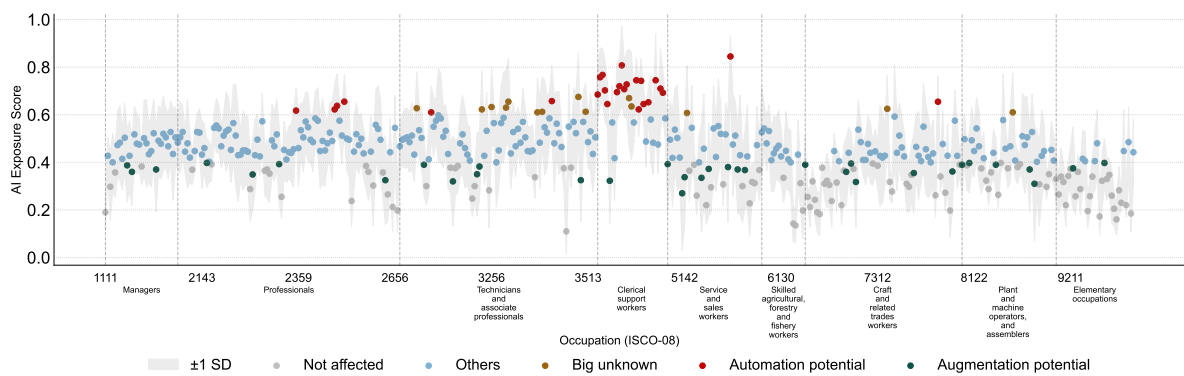
Figure 3: OE-GenAI Score and Classification



(a) low-income country



(b) middle-income country



(c) high-income country

Table 4: Occupation Exposure to GenAI with Automation Potential

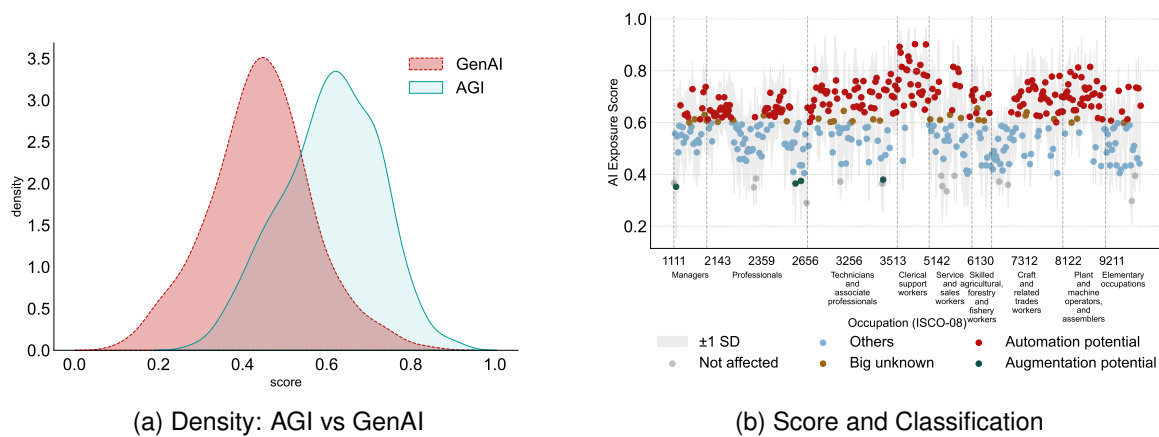
#	Occupation 1-digit	Occupation 4-digit	High-income	Middle-income	Low-income
1	Professionals	Information Technology Trainers	○		
2	Professionals	Applications Programmers	○		
3	Professionals	Software and Applications Developers	○		
4	Professionals	Computer Network Professionals	○		
5	Technicians	Incinerator and Water Treatment Operators	○		
6	Technicians	Government Licensing Officials	○	○	
7	Clerical workers	General Office Clerks	○	○	○
8	Clerical workers	Secretaries (general)	○		
9	Clerical workers	Typists and Word Processing Operators	○	○	
10	Clerical workers	Data Entry Clerks	○	○	○
11	Clerical workers	Bank Tellers	○	○	
12	Clerical workers	Travel Consultants	○	○	○
13	Clerical workers	Contact Centre Clerks	○	○	○
14	Clerical workers	Telephone Operators	○	○	○
15	Clerical workers	Hotel Receptionists	○	○	○
16	Clerical workers	Inquiry Clerks	○	○	○
17	Clerical workers	Accounting Clerks	○	○	○
18	Clerical workers	Finance and Insurance Clerks	○	○	○
19	Clerical workers	Payroll Clerks	○	○	
20	Clerical workers	Stock Clerks	○		
21	Clerical workers	Transport Clerks	○	○	○
22	Clerical workers	Proofreading Clerks	○	○	○
23	Clerical workers	Filing Clerks	○		
24	Clerical workers	Personnel Clerks	○	○	○
25	Sales workers	Contact Centre Salespersons	○	○	○
26	Trades workers	Garment Patternmakers	○		

2.4 Beyond GenAI

The development of the AI landscape has been rapidly accelerating, fueled by improvements in hardware infrastructure, breakthroughs in machine learning algorithms, and the growing availability of large-scale datasets. A long-standing aspiration within this space is the realization of Artificial General Intelligence (AGI)—an advanced form of AI capable of understanding, learning, and applying knowledge across a wide range of tasks, much like human cognitive ability (Figure 1). While AGI remains elusive, it is no longer viewed as an impossibility, especially amidst fast improvements in GenAI (Gao et al. 2024). The timeline for its emergence remains a topic of active debate, with expert predictions varying widely—from a few years to several decades. A recent global survey of over 8,000 scientists suggests a median estimate for the arrival of AGI around 2040, notably earlier than the 2060 projection reported just a few years ago (Dilmegani and Ermut 2025). However, other studies indicate an even more imminent timeline, placing the arrival of AGI within this decade (Korinek 2024). On the other hand, in 2025, the latest tests conducted by Jones and Bergen (2025) demonstrate that large language models (LLMs) have passed the Turing Test, marking significant progress in benchmarking machine intelligence against human capabilities. Prompts for AGI scoring were adapted from the GenAI ones, emphasizing hypothetical general capabilities (see Appendix for details).

As shown in Figure 4, the potential impact of AGI is substantial. The average occupational exposure score exceeds 0.6, compared to 0.43 for current-generation GenAI. Moreover, a significantly larger share of occupations is classified as automation-prone, highlighting the broader and deeper economic implications that AGI may bring once realized.

Figure 4: OE-AGI Occupation Risks



3 A Country Case

This section leverages the TE-GenAI and OE-GenAI evaluation frameworks alongside detailed labor employment data to assess Brunei’s workforce exposure to GenAI. Brunei, a high-income, oil-producing small economy in Southeast Asia, serves as a country case for this analysis. The task exposure scores were based on the setting of a high-income country, as shown in Subsection 2.3. Meanwhile, a separate exercise with in-context fine-tuning the country context to Brunei and discussions are provided in Appendix A.4.

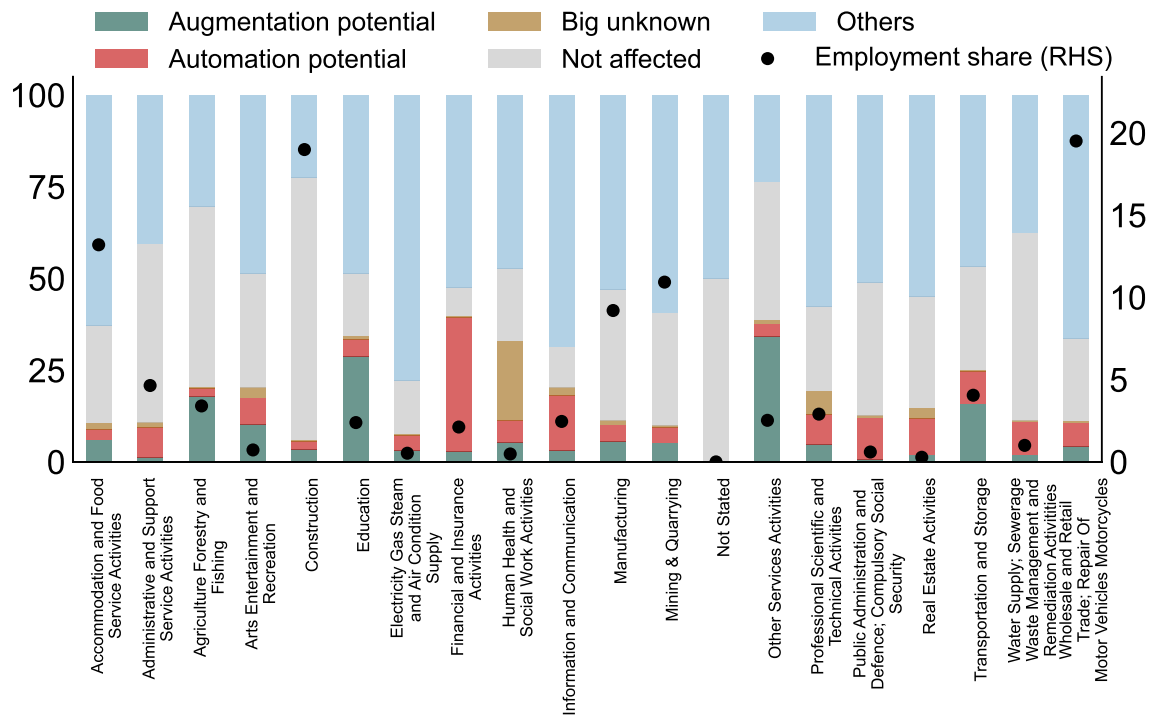
The labor data were sourced from the Department of Economic Planning and Statistics under the Bruneian Ministry of Finance and Economy (DEPS), based on the Employer and Employee Census 2023. These data are disaggregated by economic sector, gender, education level, and ISCO-08 4-digit occupation codes. The dataset covers private sector employment, excluding private household services (e.g., domestic helpers), and provides a comprehensive snapshot of Brunei’s labor market. However, it is important to note that the public sector—which accounts for approximately one-third of total employment—is excluded from the analysis. This omission may bias estimates downward, as public roles often involve routine tasks with potentially high exposure, though buffered by institutional norms.

The OE-GenAI scores were mapped to the employment data via ISCO-08 4-digit codes. By integrating these two datasets, this analysis offers a nuanced view of how GenAI’s transformative potential intersects with Brunei’s labor market structure, highlighting both vulnerabilities and opportunities across a diverse occupational landscape.

3.1 OE-GenAI in Brunei

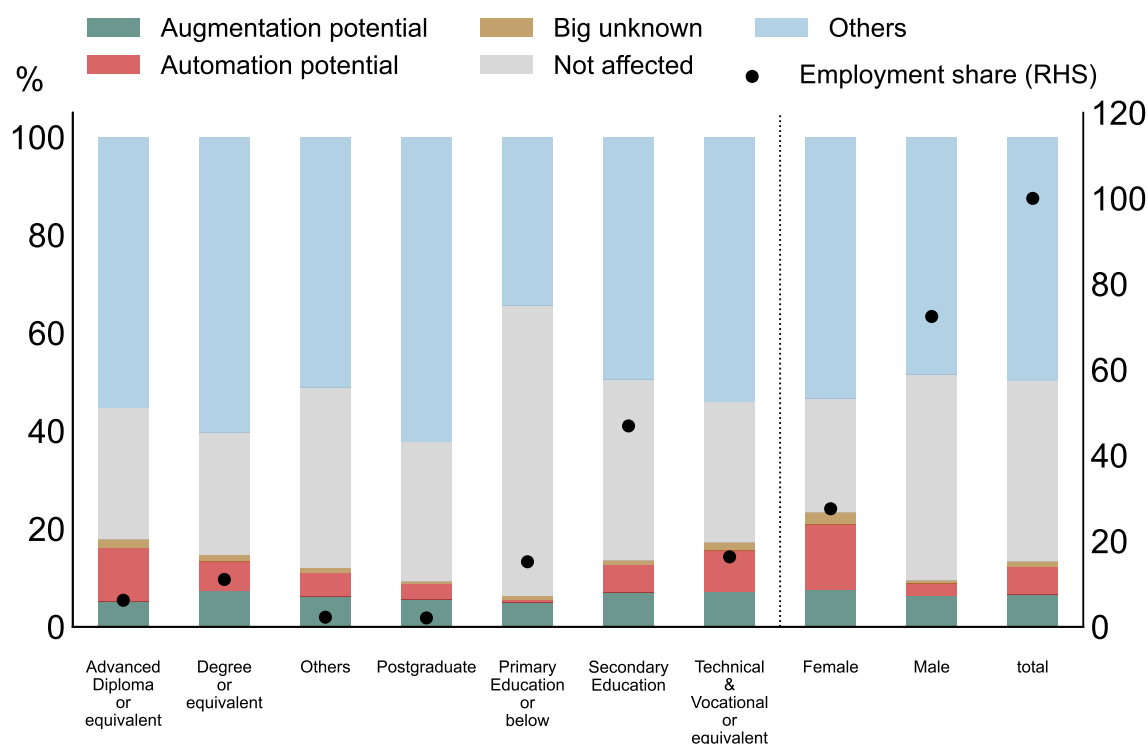
Based on the labor employment data, Figure 5 (right-hand axis) shows that Brunei’s primary employment sectors—wholesale and retail trade, construction, accommodation and food services, mining and quarrying, and manufacturing—collectively account for 72% of total employment. The same chart (left-hand axis) suggests that sectors with higher automation potential include financial and insurance activities, followed by information and communications, government-related functions (public administration, defense, and compulsory social security), and professional, scientific, and technical services. These domains are characterized by structured, rule-based tasks that GenAI is well-suited to automate, such as data processing and compliance reporting. In contrast, sectors with strong augmentation potential—where GenAI enhances rather than replaces human input—include education, other service activities, agriculture, and transport and storage. In these areas, GenAI could support creative problem-solving, pedagogical innovation, and potentially logistical coordination, underscoring its complementary role in enhancing human capabilities.

Figure 5: OE-GenAI Classification, by Economic Sector



Further, as shown in Figure 6, the disaggregation by education level reveals that individuals with mid-tier qualifications—such as secondary education, advanced diplomas, undergraduate degrees, and vocational or technical training—are more exposed to both automation and augmentation. In contrast, workers with either primary education or higher education degrees exhibit lower exposure. This pattern likely reflects the routine, codifiable nature of many mid-level tasks, which are more susceptible to automation, versus the less structured and more cognitively demanding nature of high-skilled roles.

Figure 6: OE-GenAI Classification, by Education and Gender



Gender-based analysis indicates that female workers in Brunei are significantly more exposed to automation risk than their male counterparts: 13.5% of female employment is exposed to automation potential compared to only 2.6% of male employment. This gap mirrors global trends, although the disparity in Brunei appears to be more pronounced. In contrast, augmentation potential shows a much smaller gender difference, suggesting that the opportunity for AI to complement human work is more evenly distributed across genders.

Overall, 5.6% of Brunei's private-sector employment is at risk of automation, while 6.6% shows potential for augmentation. These figures are broadly consistent with estimates from [Gmyrek, Berg, and Bescond \(2023\)](#), who focus on GPT technologies and find that in high-income countries globally, 5.1% of jobs are prone to automation and 13.4% to augmentation, with women experiencing twice the level of automation exposure compared to men. Differences may stem from Brunei's economic structure and exclusion of the public sector. In terms of shares, [Handa et al. \(2025\)](#) demonstrate that AI adoption is predominantly oriented toward augmentation (57%) rather than automation (43%). They further note that AI use is lower in both the lowest- and highest-paid roles, likely reflecting the limitations of current AI capabilities and practical barriers to implementation. The education-disaggregated view reinforces the observation that workers with mid-tier qualifications—such as

diplomas, technical, and vocational education—are more exposed than those with either lower or higher educational attainment.

A noteworthy *big unknown* category also emerges in the Bruneian context, particularly in the domain of human health and social work activities. This group comprises occupations that exhibit both high average task-level exposure scores and substantial variability across tasks, reflecting uncertainty regarding the net impact of GenAI.

4 Discussions

4.1 Implications

This paper's analysis suggests that the overall means across groups are smaller than 0.5 (with the scope between 0-1), suggesting relative mild automation potential as a whole. On the other hand, there are systematical difference in terms of GenAI exposure at the occupation task level and resulting occupation level, where jobs in high-income countries are more prone to automation potential.

Applying for a country case study, given Brunei's labor market structure and the current development of GenAI, overall exposure to GenAI remains relatively limited. On balance, augmentation potential tends to outweigh automation risk, reflecting a labor market where GenAI is more likely to complement rather than replace human work. Nonetheless, specific occupations—particularly clerical and support roles—face elevated exposure due to the nature of their tasks, such as data entry, document processing, and routine correspondence, which are highly amenable to automation or enhancement through GenAI. High-exposure sectors, including financial and insurance services as well as administrative and support service activities, are especially vulnerable given their reliance on structured, rule-based processes. In such roles—for example, insurance underwriters or paralegals—GenAI may either displace workers or significantly alter workflows toward AI-assisted models.

Looking ahead, our forward-looking predictive assessment incorporating Artificial General Intelligence (AGI) scenarios suggests that labor market's AI exposure will intensify as technology progresses. Exposure is expected to broaden across both occupations and sectors, with increases in both augmentation and automation potential. This shift may amplify risks and opportunities, suggesting a future where GenAI-driven augmentation evolves into more substantial transformation, potentially altering labor demand and job content across the economy.

Several factors will influence how exposure to GenAI and AGI translates into tangible labor market impacts leading job displacement and job restructuring. First, institutional frameworks and prevailing social norms can act as buffers, mediating the extent of disruption—even for occupations with high technical exposure ([Cazzaniga et al. 2024](#)). The magnitude of this shielding effect varies by country, depending on structural and cultural contexts. In Brunei, for instance, a possible stronger so-

cial values that prioritize collective welfare and job stability—combined with institutional preferences for face-to-face interactions in sectors like public administration—may decelerate automation in favor of preserving human-centric roles. This is particularly evident in government services, where roles such as clerks and service officers remain prevalent despite their technical exposure, partly due to the value placed on in-person engagement. By contrast, in economies dominated by private-sector market force such as the United States, if operational efficiency is more prioritized, AI adoption may proceed more rapidly, with correspondingly greater labor displacement.

Second, current measurement approaches typically overlook second-order effects and general equilibrium dynamics. For instance, within a single occupation, workers whose productivity is significantly enhanced through AI augmentation may outcompete peers with lower productivity gains, especially if labor demand does not expand proportionally in the short term. This dynamic could lead to a paradox where even occupations with high augmentation potential face displacement pressures—not through direct automation, but via intensified intra-occupational competition. In addition, as pointed out by [Acemoglu and Restrepo \(2022\)](#), the general equilibrium effect with changes in industry composition and due to task reallocation as, could lead to profound second-round effects and is beyond the scope of the current study.

Recognizing that occupations such as clerical and financial service roles are particularly vulnerable due to GenAI's ability to perform or support routine tasks, targeted policy responses are essential. Government-led upskilling and reskilling programs—such as AITI's Digital Upskilling Training in Brunei—can be refined to promote the acquisition of AI-complementary skills. Broader AI literacy initiatives across sectors will also be critical in preparing the workforce for transformation. Additionally, economic diversification efforts could align with AI adoption, promoting AI-driven productivity growth in education and creative industries, and incentivizing AI entrepreneurship in high-potential sectors. Especially, in an oil-dependent economy with a narrow economic base and large public sector like Brunei, recognizing the role of GenAI and broader AI's exposure and potential impact could help well position and prepare itself in economic and structural transformation, taking advantage of AI in improving productivity (i.e. [Brynjolfsson, Li, and Raymond 2025](#)) while managing risks.

4.2 Suggestions for labor statistics

To effectively monitor the exposure to and potential impact of AI—including GenAI and the broader AI landscape—addressing existing data gaps is a critical starting step. In the case of Brunei, the annual Labour Force Survey conducted by DEPS offers a solid platform for integrating AI-related questions to enrich the dataset.

For higher-frequency monitoring, an alternative platform could involve leveraging the survey infrastructure of the Brunei Darussalam Central Bank (BDCB). BDCB's monthly Business Sentiment Index, which surveys over 500 businesses across 11 key economic sectors, could present a more timely channel in monitoring and information gathering. AI-related questions could be incorporated

into the employment subsection of this survey to capture relevant insights.

While integrating AI questions into existing surveys offers practical advantages, a standalone AI-focused survey could also be considered. Such an approach would allow greater flexibility, including the ability to target individual workers directly and explore more specific dimensions of AI's impact on the workforce.

On the employer side, for instance, the survey design could include questions related to the current or planned adoption of AI technologies, qualitative assessments of their impact on productivity, perceived challenges, and areas of potential opportunity. An illustrative example of such a questionnaire is provided in Appendix A.6.

4.3 Limitations and Areas for Future Research

The linkages between tasks and occupations, derived through mean and standard deviation, provide valuable insights but remain inherently static. These statistical measures capture a snapshot of relationships without accounting for dynamic adjustments of tasks in relation to occupation over time. The classification threshold distinguishing automation from augmentation offers a practical framework for categorization given the constraints, yet it is fundamentally arbitrary.

The current approach facilitates exposure assessment—identifying which tasks or occupations are susceptible to generative artificial intelligence (GenAI)—but stops short of impact assessment. Translating exposure into tangible impact remains a significant challenge, as it requires integrating a broad array of variables, including institutional structures, regulatory frameworks, labor market dynamics, and technological adoption rates.

To strengthen these analyses, additional data for verification is essential. Occupation-specific case studies, coupled with a more granular understanding of how GenAI influences individual roles given a context, could enrich the broader perspective. Meanwhile, platform-based usage data, as explored by [Handa et al. \(2025\)](#), provides a practical foundation for understanding how GenAI is adopted across different occupations, despite the limitations including potential sample bias, inconsistencies in classification reliability, and incomplete datasets. These shortcomings could be mitigated if more GenAI providers disclose detailed, standardized information about usage and implementation, thereby enhancing the transparency and robustness of such studies.

Looking ahead, the increasing availability of more advanced LLMs presents a promising opportunity to advance research on labor market exposure. One potential direction involves leveraging multiple LLMs in tandem to address model uncertainty and bias. In particular, by constructing a multi-agent environment where these models collaborate—cross-validating predictions and synthesizing insights—this approach could enhance the robustness of exposure assessments. Such a framework might enable researchers to better mitigate the limitations inherent in single-model analyses.

5 Conclusions

This paper enhances a LLM-evaluation task-based framework to assess the exposure of occupations to Generative AI in broader country context and also specifically a country case, leveraging both disaggregated task evaluations and detailed labor force data. The findings suggest that country income differences do make a difference in automation exposure, with higher-income countries likely experiencing higher exposure. While overall exposure to GenAI remains modest, certain occupations—particularly in clerical, administrative, and financial services—face significant transformation potential. Augmentation currently outweighs automation risk, but this balance may shift with the advent of more advanced AI technologies such as AGI. Context-specific factors including institutional settings and cultural norms should be taken into account, in shaping the real-world impact of GenAI. Targeted workforce development, re-skilling initiatives, and adaptive social protection policies will be critical to ensuring that the gains from AI are inclusive and aligned with national development goals.

A Appendix

A.1 Models and Prompts

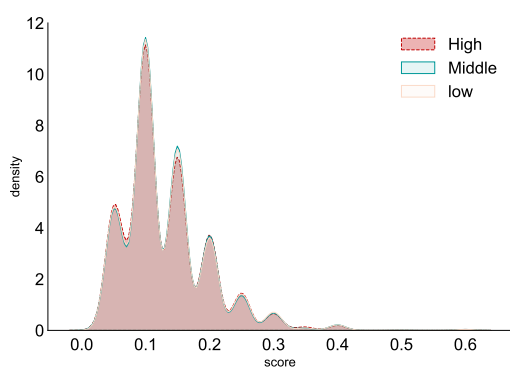
Table A1: System and User Prompts

1. System	
You are a skills specialist with deep knowledge of AI technologies—including machine learning, deep learning, GPT, and other GenAI technologies. You understand how these systems operate, their strengths and limitations, and you assess the skills and time required for specific tasks to evaluate their automation potential. Additionally, you offer insights on how AI can replace or augment human roles and advise on how individuals can adapt or acquire new skills to remain competitive in an automated job market.	
2. User	
Topic	Generative AI (GenAI); Artificial General Intelligence (AGI)
Country	A chosen country setting
Prompts	msg1 + msg2 + msg3 + msg4 + msg_con + msg_tip + msg_cot
msg1	Look at the job tasks related to occupation {title_ } with ISCO code {isco_08 }: task_list = "\n".join([f"{idx + 1}) {task}" for idx, task in enumerate(tasks)])
msg2	The job is located in {country}. Carefully assess the potential for automation of this task using {topic}. The score should range from 0 to 1 (with two decimal places), where a higher score indicates a greater potential for automation. Your reasoning should explicitly consider the capabilities and limitations of {topic} technologies.
msg3	For each task associated with a job, estimate the proportion of work time spent on that task. The value should be between 0 and 1, ensuring that the sum of all task allocations equals exactly 1.
msg4	Provide your response strictly in JSON format. Use the following structure: Outer keys should represent task numbers, and inner keys should include 'score', 'time_allocation', and 'justification'. Do not include any additional commentary.

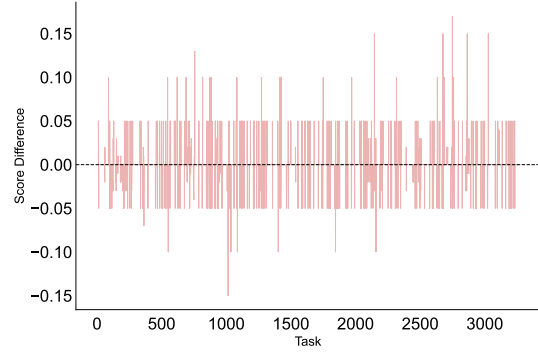
msg_con	Carefully analyze each task before assigning a score, considering both automation feasibility and current technological limitations . Your scoring logic must be consistent: If a task receives a score below 0.5, the justification should emphasize the challenges and constraints preventing automation. The justification should clearly illustrate how {topic} can replace or assist in performing the task.
msg_tip	You will receive a US\$1000 reward for a high-quality, well-reasoned response but will be fined US\$1000 for poor analysis. Ensure rigorous logical reasoning.
msg_cot	Let's not only think step by step but also analyze each aspect of the task sequentially, considering different dimensions before making a final judgment. In your justification, explicitly outline your reasoning based only on {topic} and in the context of {country}. Follow a detailed Chain of Thought (CoT) approach, systematically considering: - The task's complexity and structure. - The ability of {topic} to fully or partially perform the task. - Limitations such as contextual understanding, real-time adaptability, and interaction requirements. - Any uncertainties or constraints that impact automation potential. Each justification must be structured logically and transparently.

A.2 Additional Information of TE-GenAI

Figure A1: Statistics of TE-GenAI, Time Allocation



(a) Density



(b) Difference: High vs Middle Income Country, Time Allocation

Table A2: Paired T-Test for Task Time Allocation

Pair	T-statistic
High vs. Middle Income	-0.09 [°]
High vs. Low Income	-0.46 [°]
Middle vs. Low Income	-0.41 [°]

Note: [°] :> 0.1; *:< 0.1; **: < 0.05; ***:< 0.01.

Table A3: Occupation Task Exposure to GenAI, Highes2t Group (score=0.95)

Country Group	ISCO code	Occupation	Task	Justification
High-income	4223	Telephone Switchboard Operators	Operating switchboards and consoles to connect, hold, transfer, and disconnect telephone calls	Task involves routine operations suitable for automation via GenAI using voice recognition and NLP. High potential for automation, though real-time adaptability remains a challenge. In high-income countries, infrastructure supports automation.
	5244	Contact Centre Salespersons	Recording notes for follow-up action and updating marketing databases	Task is highly structured, with minimal decision-making. GenAI can automate data processing efficiently. High-income country infrastructure supports full automation.
High-income	5244	Contact Centre Salespersons	Maintaining statistics of calls made and successes achieved	Structured and rule-based task. GenAI can efficiently track and update statistics, with full automation possible in high-income countries.
Middle-income	4221	Travel Consultants and Clerks	Con- Issuing tickets, boarding passes, and vouchers	GenAI can generate and distribute travel documents efficiently. Limitation: Handling customer queries and last-minute changes may require human intervention. Infrastructure reliability may vary in middle-income countries.

Continued on next page

Continued from previous page

Country Group	ISCO code	Occupation	Task	Justification
Middle-income	4223	Telephone Switchboard Operators	Operating switchboards and consoles to connect, hold, transfer, and disconnect telephone calls	GenAI can fully automate this structured task using NLP. Automation potential is high, though infrastructure constraints may exist in middle-income countries.
Middle-income	5244	Contact Centre Salespersons	Maintaining statistics of calls made and successes achieved	A structured data management task that GenAI can automate almost entirely. Data availability might be a constraint in middle-income countries.
Low-income	5244	Contact Centre Salespersons	Maintaining statistics of calls made and successes achieved	Highly structured task, ideal for automation. In low-income countries, data accuracy might be a concern, but GenAI can fully replace human effort.

Table A4: Occupation Task Exposure to GenAI, Lowest Group (score=0)

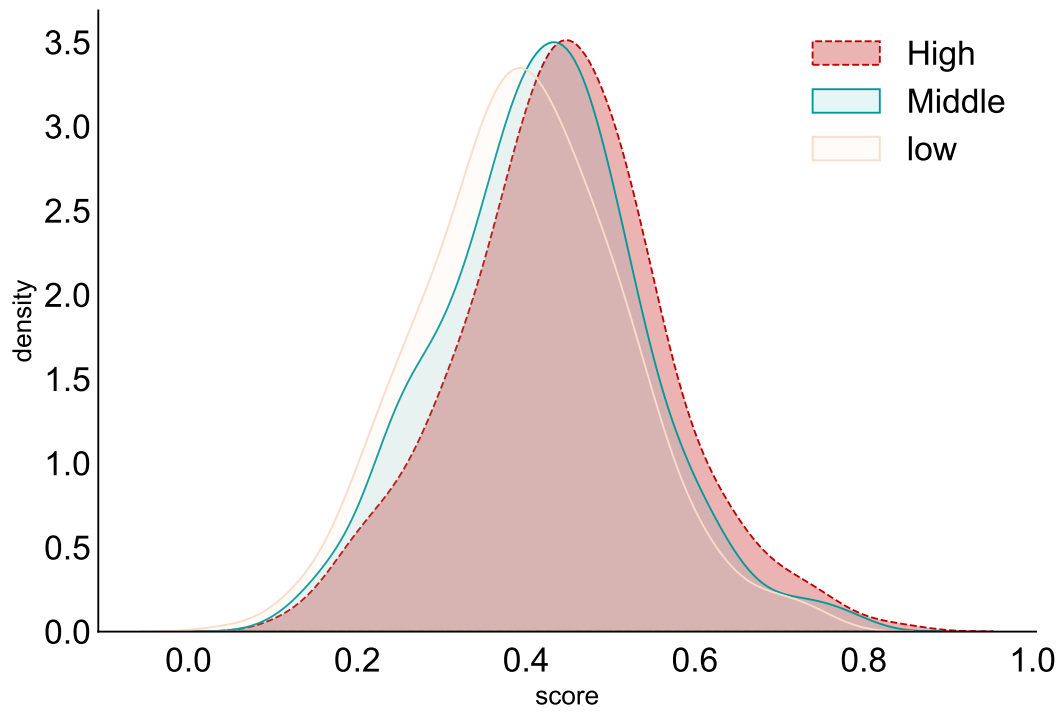
Country Group	ISCO code	Occupation	Task	Justification
High-income	3421	Athletes and Sports Players	Participating in competitive sporting events	Requires physical presence and real-time decision-making that GenAI cannot replicate. The task involves physical exertion, strategic gameplay, and immediate responses, which are beyond GenAI's capabilities. Human interaction and adaptability remain essential.
High-income	3421	Athletes and Sports Players	Competing in sporting events	Similar to participation, competing requires physicality, real-time strategy, and engagement with dynamic conditions, making automation infeasible. GenAI cannot replicate physical execution or decision-making under pressure.
Middle-income	3421	Athletes and Sports Players	Participating in competitive sporting events	Requires physical activity and real-time adaptability, which GenAI cannot perform. GenAI lacks the ability to interact physically and make spontaneous decisions. The nature of sports demands human presence and response.
Middle-income	3421	Athletes and Sports Players	Competing in sporting events	Competing in sports involves physical execution and strategic adaptation, which are beyond GenAI's scope. The ability to analyze performance exists, but real-time play cannot be automated.
Low-income	3421	Athletes and Sports Players	Participating in competitive sporting events	Requires motor skills, real-time adaptability, and strategic interaction, which GenAI cannot replace. In low-income settings, AI infrastructure limitations make even analysis support less viable.
Low-income	3421	Athletes and Sports Players	Competing in sporting events	GenAI can assist in performance analysis but cannot replace physical execution and strategic interaction. The task remains entirely dependent on human capabilities.

Table A5: Example of AI Exposure Scores and Justifications: Domestic Housekeepers (5152), "High-Income Country" vs "Brunei"

Task	Score: High-Income	Score: Brunei	Justification (High-Income)	Justification (Brunei)
Supervising workers employed in households as domestic staff	0.25	0.25	Supervising workers involves complex human interactions and decision-making that GenAI struggles to replicate. GenAI can assist with scheduling but lacks contextual understanding, adaptability, and ethical judgment.	GenAI can assist in scheduling and monitoring tasks but cannot handle interpersonal complexity. Effectiveness depends on household dynamics; the task remains human-centric even in Brunei.
Purchasing or controlling the purchase of supplies	0.65	0.65	GenAI can help analyze inventory and predict needs but lacks skills like negotiation and trend analysis. Human oversight remains necessary.	GenAI can automate ordering processes but may struggle with preferences and sudden changes. Infrastructure in Brunei supports this but oversight is still needed.
Controlling storage and issue of supplies	0.70	0.70	GenAI can automate inventory management and issuance but cannot handle unexpected issues or physical tasks.	GenAI can track inventory and issue supplies. Still needs human intervention for changing needs. Brunei likely has the tech, but human oversight is essential.
Assisting in minor injuries or illness	0.10	0.10	GenAI can provide information but not perform physical or empathetic medical care. Limits include real-time adaptability.	GenAI can offer basic first-aid advice but lacks empathy and physical capabilities. In Brunei, tech for monitoring exists but the task resists automation.
Sweeping or vacuum-cleaning, washing, polishing	0.30	0.80	GenAI can schedule tasks but cannot perform physical actions. Robotics limitations constrain automation.	Robotic systems can automate these tasks in Brunei, though issues remain with delicate handling and adaptability.
Making beds, cleaning bathrooms, supplying towels, soap	0.35	0.75	Physical nature and adaptability make full automation hard. GenAI helps with reminders, but not execution.	Robotics and GenAI can automate much of the task in Brunei. Still struggles with varied layouts and delicate handling.
Taking care of pets/plants, receiving visitors, answering phones	0.40	0.30	Task includes social and physical elements GenAI cannot handle well. Real-time adaptability and empathy are limits.	GenAI helps with reminders and phone tasks, but physical and interpersonal parts remain human-centered.
Preparing/cooking meals, setting/clearing tables, serving food	0.30	0.40	GenAI can assist with recipes and automate parts, but creativity and real-time adaptability are limiting.	Tech for automated cooking exists in Brunei, but full task complexity limits GenAI.
Cleaning kitchens, dishwashing	0.45	0.85	GenAI can schedule and monitor, and dishwashing machines help. Still, physical adaptability is a constraint.	Robotic systems can perform many cleaning tasks efficiently in Brunei. Limitation lies in kitchen layout adaptation and fragility handling.

A.3 Additional Information of OE-GenAI

Figure A2: Statistics of OE-GenAI: Density



A.4 A Experiment of Country Specification in Prompt

One of the key questions arising is whether the LLM is capable in differentiation of scores, with that prompts being further specified to provide country specification. To answer it, the two experiments based on the adjustments of prompts—one only mentioning a high-income country and the other specifically mentioning a country of "Brunei, a high-income country" are compared. The results show the two exercises generate very similar sets of results as shown in Figure A3. However, some differences in are also noticeable. For instance, there are four occupation's with the differences larger than 0.15 in an absolute terms.⁵ Based on the LLM's more granular differentiation on justifying its evaluation scores and manual checking through text data, it shows when a specific country prompt such as "Brunei" is specified, the LLM is not able to follow the instruction of tightly. For instance, the scores start to refer to to other technologies such as "robotics" and insert tone of more uncertainty.

Figure A3: OE-GenAI: "Brunei, a high-income country" vs "a high-income country"

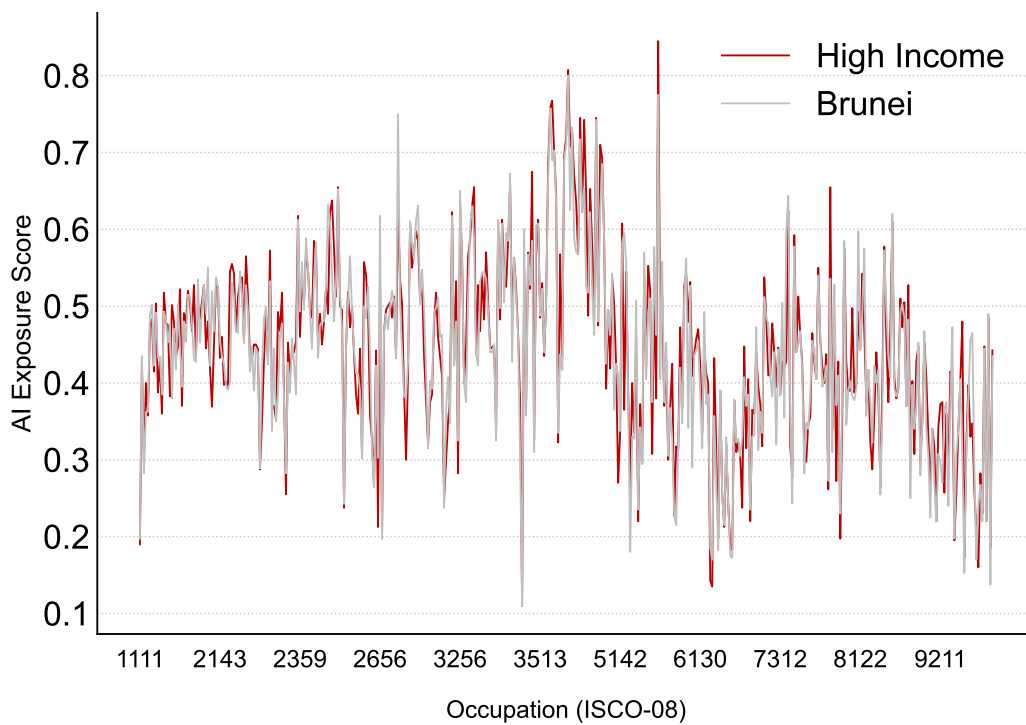
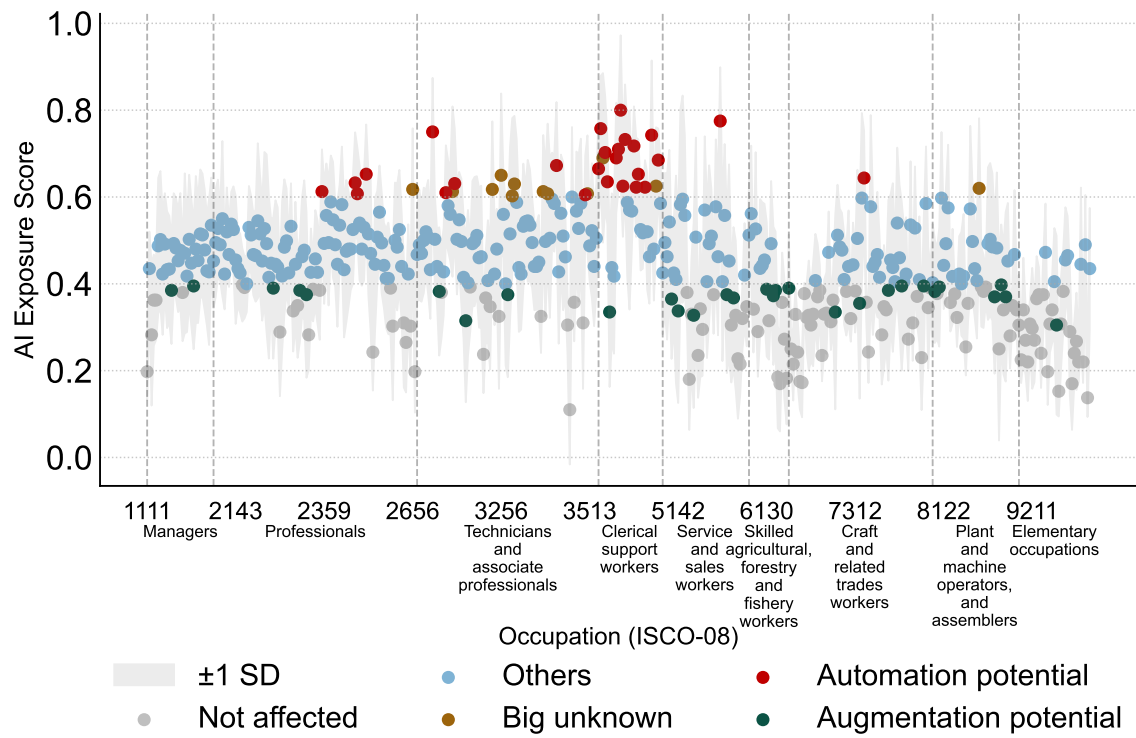


Figure A4 provides details for such occupation exposure classification, where the occupations related to clerical support workers are subject to automation potential, but also including a few occupations related to 'technicians and associated professionals, as well as professionals'.

5. It includes 2132 Farming, Forestry and Fisheries Advisers, 5152 Domestic Housekeepers, 7315 Glass Makers, Cutters, Grinders and Finishers, 9331 Hand and Pedal Vehicle Drivers.

Figure A4: OE-GenAI Classification, based on "Brunei, a high-income country"



A.5 Understanding of GPT, GenAI and AGI by Grok

A. GPT

A Generative Pre-trained Transformer (GPT) is a type of artificial intelligence model that falls under the broader category of transformer-based language models. Here's a detailed breakdown of what GPT entails:

Core Concepts:

1. **Generative:** GPT models are designed to generate text. They can produce coherent and contextually relevant sequences of words, making them useful for tasks like writing, translation, and answering questions.
2. **Pre-trained:** Before being used for specific tasks, GPT models are trained on vast amounts of text data. This pre-training phase allows the model to learn the statistical properties of language, including grammar, syntax, and even some semantic understanding.
3. **Transformer:** The architecture of GPT is based on the transformer model, which was introduced in the paper "Attention Is All You Need" by Vaswani et al. in 2017. Transformers use self-attention mechanisms to weigh the importance of different words in a sequence, allowing for more efficient and effective processing of text.

How GPT Works:

- **Pre-training:** The model is initially trained on a large corpus of text, such as books, articles, and websites. During this phase, it learns to predict the next word in a sequence, which helps it understand language patterns.
- **Fine-tuning:** After pre-training, the model can be fine-tuned on a smaller, task-specific dataset. This step adapts the model to perform well on specific applications, such as summarization, translation, or question-answering.

Key Features:

- **Scalability:** GPT models can be scaled up by increasing the number of parameters, leading to improved performance. For example, GPT-3 has 175 billion parameters, making it one of the largest language models.
- **Versatility:** Due to their pre-training on diverse texts, GPT models can be adapted to a wide range of natural language processing tasks without extensive task-specific training.
- **Zero-shot and Few-shot Learning:** GPT models can perform tasks they were not explicitly trained on, using prompts or a few examples. This capability is particularly pronounced in larger models like GPT-3.

Applications:

- **Focus on Human Skills:** Emphasizing skills that AI cannot easily replicate, such as critical thinking, emotional intelligence, and creativity, will be important for job security and advancement.
- In summary, GPT is a powerful tool in the AI landscape, capable of transforming how we interact with and generate text. Understanding its capabilities and limitations is key to leveraging its potential effectively.

B. GenAI

Key aspects of Generative AI include:

1. Types of Models:

- **Text Generation:** Models like GPT (Generative Pre-trained Transformer) can generate coherent and contextually relevant text. They are used in applications like chatbots, content creation, and language translation.
- **Image Generation:** Technologies like GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders) can create realistic images, from simple sketches to complex scenes.
- **Audio and Music Generation:** Models can compose music or generate speech that sounds natural.
- **Video Generation:** More advanced models can generate short video clips or even longer sequences.

2. Underlying Technologies:

- **Deep Learning:** Most GenAI models rely on neural networks, particularly those with many layers (deep neural networks).
- **Transformers:** A specific architecture that has been particularly successful in natural language processing tasks, enabling models to understand and generate text with high accuracy.
- **Adversarial Training:** Used in GANs, where two models (a generator and a discriminator) are trained simultaneously to improve the quality of generated outputs.

3. Applications:

- **Content Creation:** Automating the creation of articles, reports, and other written content.
- **Art and Design:** Assisting artists and designers in creating new works or modifying existing ones.
- **Entertainment:** Generating scripts, music, and visual effects for movies and games.
- **Personalization:** Tailoring content or recommendations to individual users based on their preferences and behaviors.

4. Strengths and Limitations:

- **Strengths:** GenAI can produce high-quality outputs quickly and at scale, often surpassing human speed and sometimes even quality in specific tasks.
 - **Limitations:** There can be issues with accuracy, bias, and the potential for generating misleading or harmful content. Additionally, the models require large amounts of data and computational resources to train effectively.
 - **Job Impact:** While some jobs may be fully automated, many roles will be augmented, requiring workers to collaborate with AI systems. For example, journalists might use AI to generate initial drafts but will still need to edit and fact-check.
 - **Skill Adaptation:** To remain competitive, individuals should focus on skills that complement AI, such as critical thinking, creativity, and emotional intelligence. Additionally, learning to work with AI tools and understanding their capabilities and limitations will be crucial.
- In summary, Generative AI is a powerful tool that can significantly enhance productivity and creativity across various industries. However, it also necessitates a shift in skills and roles, encouraging continuous learning and adaptation among the workforce.

C. AGI

Artificial General Intelligence (AGI) refers to a type of artificial intelligence that possesses the ability to understand, learn, and apply knowledge across a wide range of tasks at a level comparable to human intelligence. Unlike narrow or specialized AI, which is designed to perform specific tasks (such as playing chess, recognizing faces, or driving a car), AGI would have the flexibility and cognitive capabilities to handle any intellectual task that a human being can.

Key characteristics of AGI include:

1. **Reasoning and Problem-Solving:** AGI can reason through complex problems, make decisions, and solve them in a manner similar to humans.
2. **Learning and Adaptation:** AGI can learn from experience and adapt to new situations without being explicitly programmed for each scenario.
3. **Transfer Learning:** AGI can apply knowledge learned in one context to different, unrelated contexts.
4. **Abstract Thinking:** AGI can understand and manipulate abstract concepts and ideas.
5. **Generalization:** AGI can generalize from specific instances to broader principles and apply them in new situations.

Currently, AGI remains a theoretical concept and a long-term goal in the field of AI research. While significant advancements have been made in narrow AI, achieving AGI poses numerous technical and ethical challenges. Researchers are actively exploring various approaches, including neural networks, symbolic reasoning, and hybrid systems, to move closer to realizing AGI.

From a skills and automation perspective, the development of AGI would have profound implications:

- **Job Automation:** AGI could potentially automate a vast array of jobs across different sectors, from manual labor to highly skilled professions, necessitating a shift in workforce skills and roles.
- **Skill Adaptation:** Individuals would need to focus on developing skills that complement AGI, such as creativity, emotional intelligence, and complex problem-solving that require human judgment and ethical considerations.
- **New Opportunities:** AGI could also create new job categories and industries, requiring skills in AI ethics, governance, and advanced AI system management.

To remain competitive in a future where AGI might be realized, individuals should:

- **Continuous Learning:** Engage in lifelong learning to stay updated with the latest technological advancements and acquire new skills.
 - **Interdisciplinary Skills:** Develop a broad skill set that spans multiple disciplines, enhancing adaptability and versatility.
 - **Ethical and Social Skills:** Focus on skills that involve human-centric values, such as empathy, ethical decision-making, and social interaction, which are less likely to be automated.
- In summary, AGI represents a transformative potential in AI, and understanding its implications can help individuals and organizations prepare for a future where AI capabilities are significantly enhanced.

A.6 Examples of AI-related Survey Questionnaire

1. Timing of adoption

- Has your organization implemented any artificial intelligence (AI) systems in the workplace?
 - ☐ Yes ☐ No
- If yes, how long has your organization been using AI technologies?
 - ☐ Less than 1 year ☐ 1–2 years ☐ 3–5 years ☐ More than 5 years
- If no, are there plans to adopt or expand the use of AI technologies in the next 12 months?
 - ☐ Yes ☐ No
- How likely is it that AI will significantly change the nature of work in your organization within the next 5 years?
 - ☐ Very likely ☐ Somewhat likely ☐ Not likely

2. Areas of deployment

- In which functional areas are AI technologies currently being used? (Select all that apply)

– ☐ Production (operations)	– ☐ Human resources
– ☐ Customer service	– ☐ Marketing and sales
– ☐ Data analysis	– ☐ Other: _____
– ☐ Administrative tasks	

3. Impact on jobs and productivity

- Have AI technologies replaced any tasks that were previously performed by employees?
 - ☐ Yes ☐ No
- If yes, which tasks were replaced? _____
- How many employees were affected and in which roles? _____
- Has AI led to the creation of new roles or positions?
 - ☐ Yes ☐ No
- How has AI affected workplace productivity?
 - ☐ Increased significantly ☐ Increased moderately ☐ No change ☐ Decreased

4. Challenges and concerns

- What are the most common challenges faced in AI adoption? (Select all that apply)

– ☐ Lack of internal skills	– ☐ Employee resistance
– ☐ High costs	– ☐ Ethical/legal concerns
– ☐ Integration with existing systems	– ☐ Other: _____
– ☐ Data quality issues	
- Have employees reported concerns or resistance to AI adoption?
 - ☐ Yes ☐ No

5. Skills, training, and adaptation

- What skills are now most in demand since AI implementation? (Select all that apply)

– ☐ Data analysis	– ☐ System maintenance
– ☐ Programming	– ☐ Decision-making with AI
– ☐ AI model management	– ☐ Other: _____
- Has your organization provided training to support AI adaptation?
 - ☐ Yes ☐ No
- If yes, what forms of training were offered? (Select all that apply)

– ☐ In-person courses	– ☐ On-the-job training
– ☐ Online tutorials	– ☐ Other: _____

References

- Acemoglu, Daron, and Pascual Restrepo. 2022. "Tasks, automation, and the rise in US wage inequality." *Econometrica* 90 (5): 1973–2016.
- Aghion, Philippe, Benjamin F Jones, and Charles I Jones. 2017. Artificial intelligence and economic growth. Technical report. National Bureau of Economic Research.
- Arntz, Melanie, Terry Gregory, and Ulrich Zierahn. 2016. "The risk of automation for jobs in OECD countries: A comparative analysis."
- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*, qjae044.
- Bsharat, Sondos Mahmoud, Aidar Myrzakhan, and Zhiqiang Shen. 2023. "Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4." *arXiv preprint arXiv:2312.16171* 3.
- Bybee, Leland. 2023. "Surveying Generative AI's Economic Expectations." *arXiv preprint arXiv:2305.02823*.
- Carbonero, Francesco, Jeremy Davies, Ekkehard Ernst, Frank M Fossen, Daniel Samaan, and Alina Sorgner. 2023. "The impact of artificial intelligence on labor markets in developing countries: a new method with an illustration for Lao PDR and urban Viet Nam." *Journal of evolutionary economics* 33 (3): 707–736.
- Cazzaniga, Mauro, Ms Florence Jaumotte, Longji Li, Mr Giovanni Melina, Augustus J Panton, Carlo Pizzinelli, Emma J Rockall, and Ms Marina Mendes Tavares. 2024. *Gen-AI: Artificial intelligence and the future of work*. International Monetary Fund.
- Chen, Kaiji, and Yunhui Zhao. 2024. "Chinese Housing Market Sentiment Index." *IMF Working Papers* 2024, no. 264 (December). <https://www.imf.org/-/media/Files/Publications/WP/2024/English/wpica2024264-print-pdf.ashx>.
- Chen, Qin, Jinfeng Ge, Huaqing Xie, Xingcheng Xu, and Yanqing Yang. 2023. "Large Language Models at Work in China's Labor Market." *arXiv preprint arXiv:2308.08776*.
- Colombo, E., F. Mercurio, and M. Mezzanzanica. 2024. "Towards the terminator economy: Assessing job exposure to AI through LLMs." *arXiv*, arXiv: 2407.19204 [cs.AI]. <https://arxiv.org/abs/2407.19204>.
- Dilmegani, Cem, and Sila Ermut. 2025. *When Will Artificial General Intelligence Arrive?* Accessed: 2025-03-24. <https://research.aimultiple.com/artificial-general-intelligence-singularity-timing/#artificial-general-intelligence-timeline>.
- Eloundou, T., S. Manning, P. Mishkin, and D. Rock. 2023. "GPTs are GPTs: An early look at the labor market impact potential of large language models." *arXiv*, arXiv: 2303.10130 [cs.AI]. <https://arxiv.org/abs/2303.10130>.
- Felten, Ed, Manav Raj, and Robert Seamans. 2023. "How will language modelers like ChatGPT affect occupations and industries?" *arXiv preprint arXiv:2303.01157*.
- Felten, Edward, Manav Raj, and Robert Seamans. 2021. "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses." *Strategic Management Journal* 42 (12): 2195–2217.

- Frey, Carl Benedikt, and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114:254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>.
- Gao, Chen, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. 2024. "Large language models empowered agent-based modeling and simulation: A survey and perspectives." *Humanities and Social Sciences Communications* 11 (1): 1–24.
- Gmyrek, P., C. Lutz, and G. Newlands. 2024. "A technological construction of society: Comparing GPT-4 and human respondents for occupational evaluation in the UK." *British Journal of Industrial Relations*, <https://doi.org/10.1111/bjir.12840>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/bjir.12840>.
- Gmyrek, Pawel, Janine Berg, and David Bescond. 2023. "Generative AI and jobs: A global analysis of potential effects on job quantity and quality." *ILO working paper* 96.
- Handa, Kunal, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, et al. 2025. "Which Economic Tasks are Performed with AI? Evidence from Millions of Claude Conversations." *arXiv preprint arXiv:2503.04761*.
- ILO. 2023. *International Standard Classification of Occupations: Structure, group definitions and correspondence tables*. <https://webapps.ilo.org/ilostat-files/ISCO/newdocs-08-2021/ISCO-08/ISCO-08%20EN%20Vol%201.pdf>.
- Joe, E. T., S. D. Koneru, and C. J. Kirchhoff. 2024. "Assessing the effectiveness of GPT-4o in climate change evidence synthesis and systematic assessments: Preliminary insights." *arXiv*, <https://doi.org/10.48550/arXiv.2407.12826>. arXiv: 2407.12826 [cs.AI]. <https://doi.org/10.48550/arXiv.2407.12826>.
- Jones, Cameron R., and Benjamin K. Bergen. 2025. "Large Language Models Pass the Turing Test." *arXiv preprint arXiv:2503.23674*, <https://arxiv.org/pdf/2503.23674>.
- Korinek, Anton. 2024a. Generative AI for economic research: LLMs learn to collaborate and reason. Technical report. National Bureau of Economic Research.
- . 2024b. "The Economics of Transformative AI." Accessed: 2025-03-26, *NBER Reporter*, no. 2024 Number 4, <https://www.nber.org/reporter/2024number4/economics-transformative-ai>.
- Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science* 381 (6654): 187–192.
- Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. "The impact of ai on developer productivity: Evidence from github copilot." *arXiv preprint arXiv:2302.06590*.
- PWC. 2017. "Sizing the prize: what's the real value of AI for your business and how can you capitalise," <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>.
- Roumeliotis, K. I., N. D. Tselikas, and D. K. Nasiopoulos. 2024. "LLMs in e-commerce: a comparative analysis of GPT and LLaMA models in product review evaluation." *ScienceDirect*, <https://www.sciencedirect.com/science/article/pii/S2949719124000049>.

- Steyvers, Mark, Heliodoro Tejeda, Aakriti Kumar, Catarina Belem, Sheer Karny, Xinyue Hu, Lukas W Mayer, and Padhraic Smyth. 2025. "What large language models know and what people think they know." *Nature Machine Intelligence*, 1–11.
- Tan, T. F., K. Elangovan, L. Jin, Y. Jie, L. Yong, and J. Lim. 2024. "Fine-tuning Large Language Model (LLM) artificial intelligence chatbots in ophthalmology and LLM-based evaluation using GPT-4." *arXiv*, <https://doi.org/10.48550/arXiv.2402.10083>. arXiv: 2402.10083 [cs.AI]. <https://doi.org/10.48550/arXiv.2402.10083>.