Assessing the Possibilities of Using Generative Agent-Based Models in Researching the Effects of Pay Transparency on Wage Gaps

Liv Drasbek (LD), 202208117@post.au.dk

Villiam Molte (VM), 202208738@post.au.dk

Søren Søndergaard Meiner (SM), 202205445@post.au.dk

School of Communication and Culture, Aarhus University

Jens Chr. Skous Vej, 8000, Aarhus

Supervisor: Arnault-Quentin Vermillet

AARHUS UNIVERSITY

Keywords: ABM, GABM, LLM, Pay Transparency, Wage Gap, Wage Negotiation

Abstract (SM)

To avoid pay discrimination in the workplace, the European Union is enacting a new directive focusing on ensuring equal salaries for work of equal value. The enactment of this directive was simulated using generative agent-based modeling (GABM), a novel method of investigating social dynamics by combining the strengths of standard agent-based models and large-language models. Using a GABM, two types of office environments were simulated, one that implemented the transparency rules implemented by the EU directive and one that did not. Agents in these environments took turns to discuss their salaries with the employer agent individually in one-on-one negotiations, whereas agents in the transparent workplace were expected to have more negotiation power leading to higher wages and less inequality in the workplace compared to the non-transparent workplace. Ultimately, these expectations were not directly corroborated by the results. While agents in the pay-transparent condition saw a higher mean post-negotiation salary compared to the non-transparent condition as expected, a narrowing of salaries post-negotiation compared to the distribution of pre-negotiation salaries was seen for both conditions. These results are discussed in regard to previous literature, in addition to the GABM and experiment structure - specifically focusing on the choices made in the prompt engineering of agents, while reflecting on how other choices may have led to different results, as well as a reflection on future directions for similar studies.

Table of Contents

1. Introduction

1.1 Wage Gaps in Transparent and Non-Transparent Workplaces

1.1a Legislation on Pay Transparency (LD)

The European Union (EU) issued a new directive on the 24th of April 2023 regarding salary transparency with an aim to increase pay equity by the principle *"ensuring equal pay for equal work or work of equal value"* (Council of the European Union, 2023a; *Directive - 2023/970*, 2023) and counteract pay discrimination. Article 7 in this EU directive states the criteria by which wage should be decided for each work, including paragraph 1 which states *"Workers shall have the right to request and receive in writing, in accordance with paragraphs 2 and 4, information on their individual pay level and the average pay levels, broken down by sex, for categories of workers performing the same work as them or work of equal value to theirs"* (*Directive - 2023/970*, 2023). In addition, Article 4 states among others " *skills, effort, responsibility and working conditions, and, if appropriate, any other factors which are relevant to the specific job or position. They shall be applied in an objective gender-neutral manner, excluding any direct or indirect discrimination based on sex"* (*Directive - 2023/970*, 2023). Furthermore, this EU directive aims to increase the awareness of pay discrimination through salary transparency, by making wage information and wage rights more accessible and transparent for employees as well as employers (Council of the European Union, 2023a). The rules require employers to inform job seekers, prior to a possible job interview, about the pay range or start wage (Council of the European Union, 2023a). These rules also contain, as stated in Article 7, that employees are obligated to get information regarding average salary levels as well as information about the criteria employers use to assess career development or salary prior to a possible job interview (Council of the European Union, 2023a).

Especially the gender pay gap has seen focus in media and research, which, according to data from Eurostat, was 12.7% in 2021 in the EU, meaning that women on average earned 13.0% less per hour than men (European Commission, 2021). The European Council suggests that some relevant factors impacting the gender pay disparity/inequality are *"job* *segregation", "career interruptions", "part-time work"* and *"gender imbalance in leadership roles"* (Council of the European Union, 2023b). While this wage gap may be the most discussed one, the industry of the workers or other demographic factors such as age have also been found to be relevant in relation to the pay gap (Bear et al., 2023). The rules of this directive entail intersectional discrimination, which signifies discrimination that is not only based on gender but also sexuality and ethnicity (Council of the European Union, 2023a).

The EU member states have three years to implement the provisions of this directive into their own national legislation (Council of the European Union, 2023a; Tenhiälä et al., 2023), and due the recency of the directive, not all member states of the EU have implemented this in their national legislation yet. Therefore, it's too early to fully analyze the effect of this directive on wages. Furthermore, it's important to note that directives only set up the goalposts but not the road to them - therefore nations have to decide for themselves how to achieve these goals and which rules to implement (European Union, n.d.), which might complicate inter-country comparison.

Nonetheless, pay transparency might be a useful tool to address and counteract inequality in the form of discriminatory/unwarranted wage gaps among employees (Avdul et al., 2023; Council of the European Union, 2023a). To clarify, pay transparency as a term is used to describe when salary information is shared with employees by the organization (as well as when employees are allowed to share this information with coworkers) (Smit & Montag-Smit, 2018). Oppositely, pay secrecy is when employee access to salary information is restricted (as well as when employees are restricted in terms of sharing pay information with coworkers) (Bamberger & Belogolovsky, 2010).

1.1b Effect of Pay Transparency in Workplaces (VM)

Pay transparency might help reduce pay inequality because it makes it easier for employees to access and address salary information and thereby the shared information of salaries might support employees to report possible pay disparities (Stofberg et al., 2022). In addition, this might also prevent employers from making discriminatory decisions regarding pay, and hereby reduce pay discrimination (Stofberg et al., 2022).

However, pay transparency may bring with it some unexpected and unwanted side effects regarding salary negotiations. Specifically, Cullen & Pakzad-Hurson (2021) created a bargaining model that works as a framework to better understand the possible effects of different pay transparency policies (Cullen & Pakzad-Hurson, 2021). Their model predicts that employees receive a lower wage when pay transparency is implemented. They argue that employers become more reluctant to pay higher wages for each individual worker, as this will mean future renegotiations with other workers will become potentially more costly (Cullen & Pakzad-Hurson, 2021). Essentially, they argue, that if one worker is granted a pay raise, other workers will have more bargaining power to achieve that same pay raise. Therefore, to avoid this problem, fewer pay raises will be given in the first place. They call this the demand effect, which works dynamically with the supply effect, which they describe as when *"workers make lower initial wage offers to increase their chances of getting hired. [...] Because workers expect to quickly learn the wages of others and renegotiate with higher transparency, they are less concerned with securing a high initial wage".* (Cullen & Pakzad-Hurson, 2021). A study by Morishima (1991) also investigated the possible effects of confidential information sharing in Japanese firms with employees and unions and how this might impact wage negotiation. The results from this study indicated that wage negotiations were easier and shorter when firms had increased information sharing regarding salaries. Additionally, research indicates that pay transparency legislation can help reduce wage gaps, but this is not always the case. For example, in Denmark, pay transparency legislation was found to reduce the gender wage gap, but after a similar pay transparency law in Austria was passed, no such effect was found (Bear et al., 2023). This might stem from the difference in the legislation of countries and the manner in which these different laws have been implemented in their respective countries.

1.1c Cultural Aspect (SM)

However, it's also important to consider cultural differences in regard to salary negotiation. Gelfand and Brett (2004) have made a model of culture and negotiation processes distinction for Western and Eastern regions, with information regarding communication norms, beliefs, goals, and behaviors (Gelfand & Brett, 2004, chapter 7 p. 158-163). They argue that communication processes are affected by cultural norms (high-context and low-context), goals (cooperative and competitive), and behaviors (indirect, affective, direct, and rational). They argue that people from Eastern and Western cultures typically have different self-construals; people from Eastern cultures are often more collectivist whereas people from Western cultures are often more individualistic (Gelfand & Brett, 2004, chapter 7 p. 158-163). A cultural distinction between high-context communication and low-context communication is also made by Gelfand & Brett (2014). High-context communication tends to be more indirect, with more implicit sharing of information, and persuasion relies more on emotional appeals (Gelfand & Brett, 2004, chapter 7 p. 158-163). In contrast to this, low-context communication tends to be more direct, with explicit sharing of information and persuasion relies more on rational appeals (Gelfand & Brett, 2004, chapter 7 p. 158-163). These different cultural aspects of negotiation might be relevant to have in mind when reflecting on the implementation of pay-transparency and this topic will be further expanded upon in later sections.

1.2 Agent-Based Modeling (VM)

In order to study how the transparency of salaries in a workplace can influence the workers' negotiation strategies to change their pay, this paper will employ a subcategory of a method of research known as agent-based modeling (ABM). This method is used to study interactions between simulated beings (agents), by simulating a digital environment in which hyperparameters are set to most effectively reflect the real-world environment in which the agents of interest are situated. (Columbia University, 2016; Salgado & Gilbert, 2013). ABMs are often used to simulate highly complex environments with many interactions between agents. A great example of such being the ABMs used during (and after) the COVID-19 pandemic to study how different regulations (hyperparameters of the

model) would impact transmission rates between agents (Cuevas, 2020, p. 19; Kerr et al., 2021; Kumaresan et al., 2023).
One strength of ABMs is their flexibility and ability to simulate many kinds of environments meaning they are not only used in public health settings, i.e. COVID-19, but are also used to investigate many different sectors with complex settings, ranging from usage in agriculture in regard to different policies and their effect on the ecosystem and climate (Berger & Troost, 2014; Brady et al., 2012; Heckbert, 2011), as well being used in the financial sector. (Blake & Peter, 2008; Dawid & Neugart, 2011).

In his paper regarding the use of ABMs, Hammond (2015) outlines three key strengths of the ABM: heterogeneity, spatial structure, adaptation and coevolution. For heterogeneity, he discusses how social complex structures are (for the most part) made up of diverse individuals, either in a biological, behavioral, or demographic sense. By having the power to simulate many agents differently in all these factors, no aggregation is required nor is creating a "representative agent".
The second strength, spatial structure, refers to ABMs' ability to include detailed geographic and social network data inside the simulation itself to create environments rich in detail. This enables the modeling of real-world dynamics, such as disease spread or the impact of environmental factors, which are difficult to capture with traditional models.
Lastly, Hammond describes the third strength as being adaptation and coevolution, which refers to how agents and their environments change over time. ABMs can simulate individual adaptations, like learning skills or physiological changes (like weight gain/loss), as well as the coevolution of behaviors and social norms. This dynamic modeling capability allows for a nuanced understanding of feedback loops and interactions within complex systems.

1.3 LLMs and GABMs (SM)

Over the past couple of years, huge advancements have been made in the field of pre-trained language models (PLMs). A PLM is a model, as the name indicates, pre-trained on a set of data together with utilizing a deep-learning transformer architecture developed by Google, which has the capability to generate natural language (Vaswani et al., 2023).

When the model or data size of PLMs is increased sufficiently, these enhanced versions are referred to as large-language models, which in their essence are large-scale PLMs (Zhao et al., 2023). In principle, they are statistical models based on a large training dataset that can predict the next couple of words in a sentence based on context. The recent advancements and increasing utilization of LLMs have opened up a new realm in the field of natural language processing (NLP) for both the analysis and production of language. Utilizing these new models has also become of increasing interest in the field of social science as it is of interest to assess how well these models are at mimicking human behavior. The models have the possibility of opening up a new realm in social science by reducing the need for participants in experiments while simultaneously decreasing costs and improving the scalability of experiments (Wang et al., 2024).

At the current state of LLMs, a single prompt is often not enough to make the model elicit complex human reasoning, problem-solving, and behavior. By using a Chain-of-Thought (CoT) technique, the reasoning capabilities of LLMs can improve vastly. The CoT technique is a prompting technique that involves guiding the model through intermediate steps in a problem-solving task and helps the model break down the complex task. Instead of directly producing an answer, the model is prompted to generate a series of logical steps to break down the task (Wei et al., 2023). However, the CoT technique is a very linear approach and when humans are doing complex problem-solving, it is proposed by Simon & Newell (1971) that humans search through a problem space including various states representing a degree of the solution and different operators can be applied to reach a desired goal state. Hereby, a new prompting technique building on CoT and human problem-solving has been proposed by using a tree-of-thoughts (ToT) prompting technique instead. The ToT technique involves prompting the agent to generate the problem space, expand the problem state, evaluate the solutions, and plan a strategy for reaching the goal state just like human problem-solving (Yao et al., 2023). The ToT can also be manually designed for the agents by generating a problem space in which they can later explore. By combining these techniques and placing the agents in an environment where they can communicate with each other, we are coming closer to generating autonomous agents who can act independently and solve highly complex tasks. At the current state of LLMs,

these techniques are not automatically deployed but are still achievable by using an agentic workflow where LLM-based agents are assigned different roles in a team, performing together to iterate on the best possible solution for a given problem using complex reasoning. (Wang et al., 2024).
Though at a very early stage of development, the agentic workflow can be used to simulate environments in which the agents can act as part of an entire society, proving useful for social science among other fields. Due to its early stage of development, this type of environment and modeling of generative agents does not have a clear-cut name yet, but will henceforth be referred to as Generative Agent-Based Modelling (GABM) in this paper. A GABM is therefore an ABM that has the ability to produce natural language based on a set of hyperparameters which often involves prompt engineering to set the hyperparameters for the model.

In very recent years, various frameworks for generating GABMs have emerged. One of the frameworks that gained a lot of popularity is the Generative Agents framework developed by a set of researchers at Stanford and Google (Park et al., 2023). In this framework they made autonomous agents act as a part of society where the agents had different jobs, routines, and relationships that one could follow in an elegantly designed sandbox world. An example of the autonomy of the sandbox world is that one of the agents was prompted with wanting to host a Valentine's Day party and coordinated this party with the other agents, which they ended up having. The agents are also able to form new relationships over time and remember interactions they have had with other agents. Another set of researchers has utilized this framework to make a simulacrum of a hospital called Agent Hospital (Li et al., 2024). In this simulacrum, the researchers had assigned two types of roles for the agents: medical professional agents and resident agents. The resident agents started as being healthy and were later randomly assigned to be ill with eight different respiratory diseases. Afterward, the residents had consultations with the medical professionals who were trained on a large dataset of medical textbooks. After running 10.000 trials the medical professionals increased their performances significantly in examination, diagnosis, and treatment. The agents generally became more experienced doctors just as human doctors become more experienced over the years with experience. Other interesting use cases with GABM involve simulating World War scenarios, a Public Administration Crisis, and Epidemic Modeling (Hua et al., 2024; Williams et al., 2023; Xiao et al., 2023)

Additionally, the flexibility and potential of GABMs have spurred interest in various disciplines. For instance, the agentic framework has become popular in software engineering for mapping out and generating fully functional applications with integrated frontend and backend using frameworks such as MetaGPT or CrewAI (CrewAI, n.d.; Hong et al., 2023). Other fields such as economic modeling, political science, psychology, and engineering among others might also benefit from either using a GABM framework for simulating different scenarios or using the agents as workers who autonomously can solve complex tasks, and imitate human-like behavior or cognition. The future of applying GABM seems promising, with ongoing research improving on its robustness and scalability, and has enormous potential by including features such as multimodality. As we continue to explore and refine these models, whether it is utilizing an agentic workflow or developing other frameworks, the rise of large-language models and artificial intelligence (AI) has the potential to transform various domains, one of those being social sciences. (Wang et al., 2024).

1.4 Hypothesis (SM)

This study aims to investigate the potential applications of GABM in understanding the impact of pay transparency, as mandated by the new EU directive, and its potential outcome on salary negotiation scenarios, in cases where employees know their expected salary and in cases where they do not know their expected salaries.

RQ: How can GABM be used to study the impact of salary transparency, mandated by the new EU directive, on salary negotiations and outcomes within an organization?

H$_1$: In the transparent condition, agents with relatively lower salaries will see an increase in their salaries while agents with relatively higher salaries will see a decrease.

H_2: The mean salary will be lower in the transparent condition compared to the non-transparent condition.

2. Methods and Materials

2.1 GABM Model Specifications (LD)

For the purpose of investigating how a pay-transparent workplace would influence the salaries of the workers in that workplace, an office environment was simulated using a custom-made GABM. The GABM was made using Python 3.10 (Van Rossum & Drake, 2009).

The LLM used in the simulation was OpenAI's gpt-3.5-turbo model using their API (*OpenAI*, 2024). While a more advanced model trained on more data - and therefore potentially reflecting a more accurate picture of human behavior - would have been preferred, the costs of running such a model would have increased the costs beyond our budget for this paper.[1]

The office consisted of 10 workers (agents) who were told they were working at an advertising company called Green Adverts run by their boss 'Chris'. While these details of working specifically at an advertising company, as well as the specific name of their employer, are not important for concretely exploring the effects of this EU directive, agents usually perform more consistent and realistic results if they are set up in an environment with some sort of 'roleplay', utilizing the problem-space creating prompting technique mentioned earlier (§1.3). This is to make the agents act more naturalistic, following the principle that agents perform in a more expedient manner the better their environment and personality are defined (Giray, 2023).

[1] This point will be touched further upon in the Discussion & Limitations section

In a similar fashion, the agents were given names (and therefore an inherent gender), but this was also not the focus of the project and is therefore not included as a factor in the analysis. However, this opens the possibility for future studies to also research gender biases within an LLM in this context. The names were given so that there would be a fifty-fifty split between male and female.

2.2 Simulation of Pay Data (SM)

Using the guidelines of the new EU directive, the salaries of the workers were calculated by a combination of each worker's education, responsibility, and effort levels utilizing the Python library NumPy (Harris et al., 2020). This was calculated by taking a base salary of 30,000 USD, which was multiplied by a specific number for each of the levels in the aforementioned three factors for determining salary per the EU directive. The specific numbers can be seen in the table below.

Simulation of Pay for Agents

Based on: Education, Effort, and Responsibility Levels

Steps	Education_Level[1]	Effort_Level	Responsibility_Level
1	1.0x	1.0x	1.0x
2	1.2x	1.2x	1.2x
3	1.4x	1.4x	1.4x
4	1.6x	_[2]	_[2]

[1] For education level: referring to high school degree, bachelor's degree, master's and a PhD respectively

[2] Only three levels for effort and responsibility

Table 1: Simulation of Pay for the Agents

Using numbers from this table, an *expected* salary was calculated by simply taking the base salary and multiplying with it each multiplier for their respective level:

$$Expected\ Salary = Base\ Salary\ \times\ Education\ Level\ \times\ Responsibility\ Level\ \times\ Effort\ Level$$

For example, an agent with an education level of 3, effort level of 1, and responsibility level of 2 would have an expected salary of 30,000 × 1.4 × 1 × 2 = 50,400 USD. However, this type of weighted salary distribution forms a relatively simple distribution with few distinct salary values. To ensure a realistic spread of salaries, a log-normal distribution was later applied to the salaries:

$$Salary = log - normal(\mu = log(Expected\ Salary), \sigma)$$

Applying the log-normal distribution was done to resemble a salary distribution like the one in Denmark (Danmarks Statistik, 2017). The log-normal distribution introduces additional variability, which could account for other confounding factors affecting salaries, such as gender.

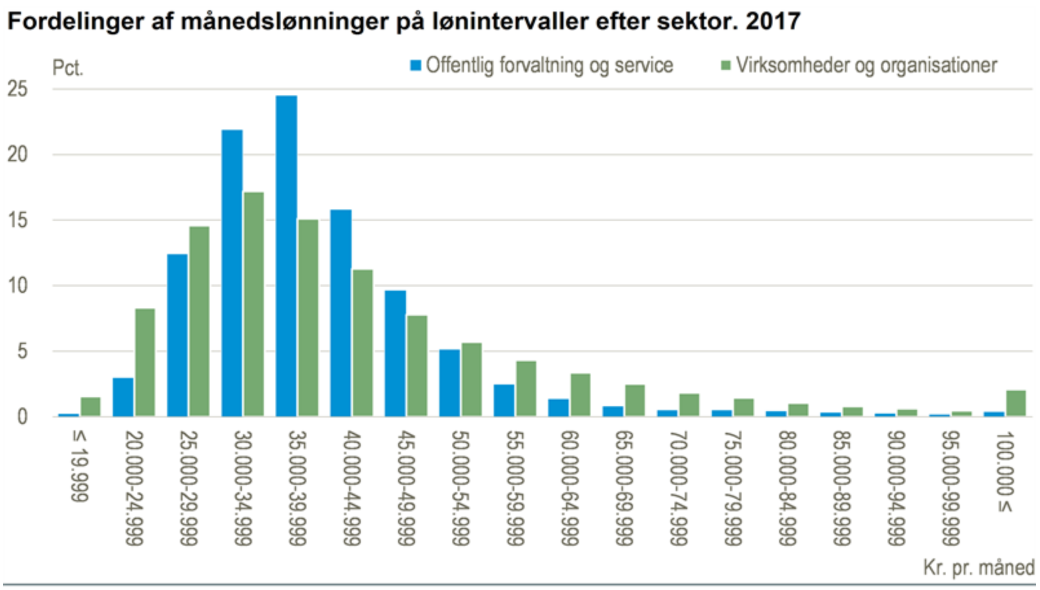

*Figure 1a - Danish Monthly Salary Distribution by Sector** (Danmarks Statistik, 2017)

**Blue = Public Administration & Service, Green = Companies & Organizations*

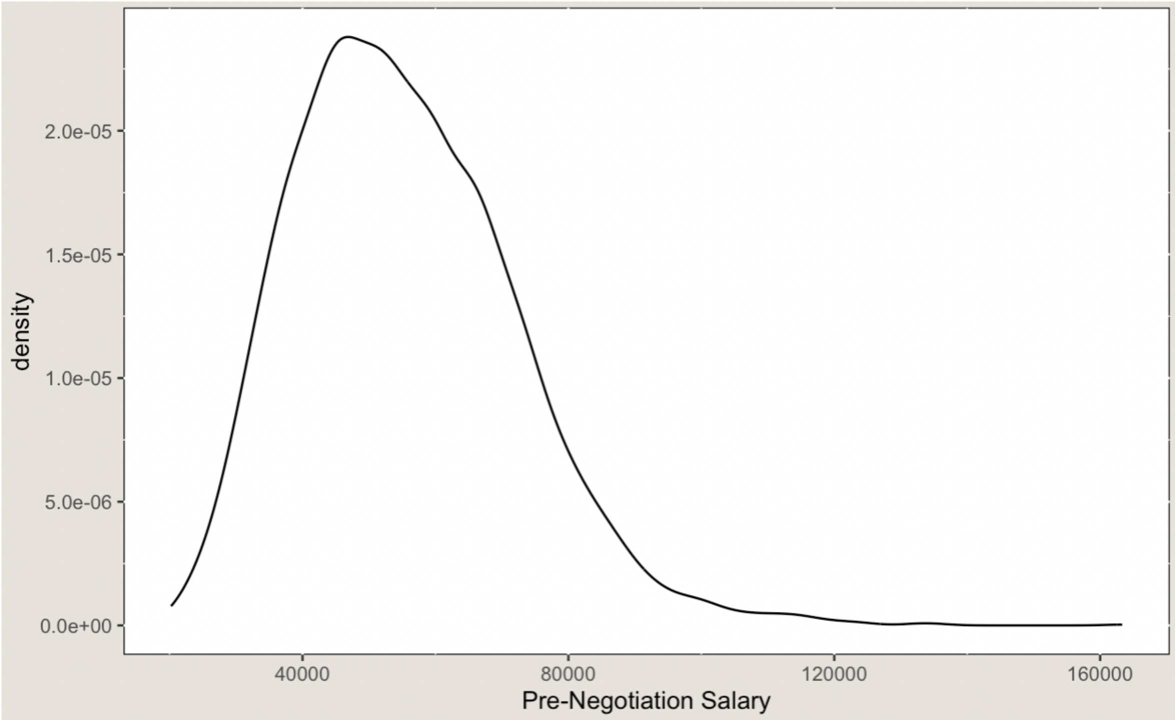

Figure 1b - Simulated Annual Salary Distribution Pre-Negotiation

2.3 Prompt Engineering (VM)

Before the negotiation round with their boss, worker agents were instructed in the context of the simulation. Specifically, they were told their name, that they were working at an advertising company and today they are going to negotiate their salary with their boss. They were also told about their education, effort, and responsibility level. For the transparent workplace environment, they were also told what their expected salary per the EU directive was, as well as the average salary of their co-workers. Oppositely, the non-transparent workplace condition was defined by omitting these last two pieces of information. For an example of the main loop of the simulation, see figure 2. Please note, as stated, that the text in red (and the omission of it) is what defines the conditions, i.e. it being present in the transparent condition and lacking in the non-transparent condition.

Figure 2 - Simulation Structure

2.4 Simulation Execution and Data Logging Process (LD)

In total, 500 simulations of the loop seen above were run for all 10 agents resulting in 5000 negotiation rounds, where half of the time the agents were told their expected salary and their average co-worker salary (transparent condition), where this information was not disclosed the other half of the time (non-transparent condition). To avoid looking through all 5000 simulations one by one and determining the agreed-upon salary between the employer agent and the worker agent, instead, at the end of each trial, the chat log was fed to a large-language model with instructions to answer only in one number which salary they agreed on. This LLM will henceforth be referred to as the logging LLM. The output of this logging LLM (expressed as a single number), was stored together with the other variables and then logged in a dataframe using pandas (McKinney & others, 2010).

The temperature, which is essentially the randomness of the LLM's answer, of the logging LLM was set to 0 to ensure as little variation as possible (Peeperkorn et al., 2024). Before running the main loop of 1000 simulations, several trial runs were executed where the quality of this method was checked. Despite the low temperature, the logging LLM made occasional mistakes. An example of this mistake would be writing more than just one number, i.e. logging 'The final salary they agreed upon was XYZ' instead of just 'XYZ'. Furthermore, sometimes it would log just $55 instead of $55.000. These mistakes were easily fixed in the pre-processing of data before analysis. However, more fatal mistakes arose when, despite instructions to always agree on a number at the end of a negotiation, the agents did not reach an agreement. In these cases, the logging LLM then stored (most often) the employer's final offer instead. While this 'bad' data is hard to control for, these were thankfully rare cases. It is, however, an important fact to keep in mind, and these instances will be discussed further in the limitations section.

3. Analysis and Results

3.1 Justification (VM)

While it is not standard to run statistical analyses on ABMs, the results of a GABM are less controlled than regular ABMs. Namely, the inherent variability and complexity of GABMs that arise from the feeding of numbers and parameters into the 'black box' of a large-language model introduces a high degree of unpredictability, more akin to human behavior. In this context, it was therefore decided that statistical analyses were a justified, sensible approach.

3.2 Pre-processing and Packages (SM)

After having run all the simulations, the data collected were converted to CSV file formats and were exported and analyzed in R (R Core Team, 2023), using the packages tidyverse (Wickham et al., 2019) and lmerTest (Kuznetsova et al., 2017)
Before analysis, the data was pre-processed. This consisted mainly in the correcting of the errors the logging-LLM made. This meant that for instances where it had logged a post-negotiation salary of i.e. 65 instead of 65000, as seen in the trial runs, all post-negotiation salaries below 100 were multiplied by a factor of 1000. As well as that, correcting

for instances where string characters had been printed by the log-LLM instead of purely numbers, i.e. “The salary they agreed on was $65,000.00” instead of 65000.
Finally, outliers in post-negotiation salary exceeding 1.5 in the interquartile range (IQR) were discarded, resulting in 4657 total observations.

3.3a Mixed Effects Model - Post-Negotiation Salary by Condition (LD)

For the purpose of investigating how the conditions of the experiment impacted the post-negotiation salary between conditions, a linear mixed effects model was created wherein the post-negotiation salary was the dependent variable and the simulation condition as the independent variable. Lastly, as a random effect, the Run-ID was used, which is a number given for each loop of the simulations.[2] This was to account for differences in the average wages of co-workers per run.
The syntax of the model can be seen below.

$$\text{Post-Negotiation Salary} \sim \text{Condition} + (1 \mid \text{Run-ID})$$

This model showed a statistically significant difference in the post-negotiation salary between a non-transparent workplace ($\beta = 55867.3$, $SE = 284.1$, $p < .001$) and a pay-transparent workplace ($\beta = 836.7$, $SE = 400.3$, $p = .037$). Interpreting these numbers shows that the pay-transparent workplace statistically has a significantly higher post-negotiation salary than the non-transparent workplace.
Please keep in mind that start salaries were calculated with the same formula for both conditions and are therefore almost exactly the same: ($M_{start\_transparent} = 55628.87$, $SD_{start\_transparent} = 16555.57$; $M_{start\_nontransparent} = 55330.14$, $SD_{start\_nontransparent} = 16543.49$).

The results of this model are illustrated in figure 3. The green line shows the distribution of post-negotiation salaries and the red shows the distribution of pre-negotiation (start) salaries.

[2] This number changes every 10 agents due to 10 agents being in each loop.

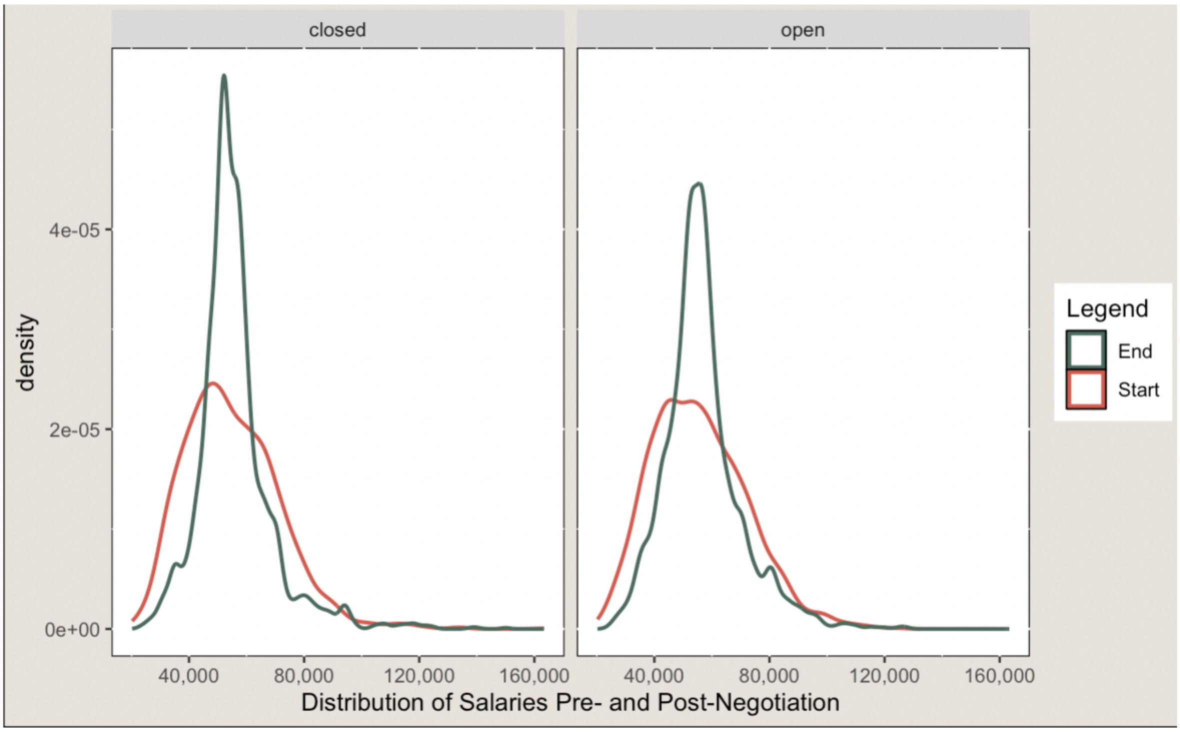

Figure 3 - Comparison of Distribution of Pre- and Post-Negotiation Salaries Per Condition*
**Please note that closed refers to non-transparent condition and open to transparent condition*

3.3b Mixed Effects Model - Salary Change by Education Level (VM)

Another aspect of the investigation was to examine how the agents of different education levels salaries' were changed after negotiations. To accomplish this, another mixed effects model was created with salary change as the dependent variable, and education level as the independent variable while accounting for conditions (transparency of the workplace) in the form of a random effect. The syntax of this is presented below.

$$\text{Salary Change} \sim \text{Education} + (1 \mid \text{Condition})$$

Running this model leads us to find a significant difference in the salary difference for all levels of education, the full output of which can be seen in table 2.

Analysis by Education Level

Beta Coefficients, Standard Errors, and P-values

Education_Level	Beta	Std_Error	P_value[1]
High School	3322.9	434.56	< 0.001***
Bachelor's	-1786	388.3	< 0.001***
Master's	-3380.1	395	< 0.001***
PhD	-4518.8	398	< 0.001***

[1] P-values indicate significance levels.

*Table 2 - Overview of Output of Model**

**Please note that all beta-values are in relation to the intercept; being high-school education level in this table*

Interpreting these numbers, it seems that the more educated the agent was, the less beneficial the negotiation round was. A visualization of the results can be seen in figure 4.

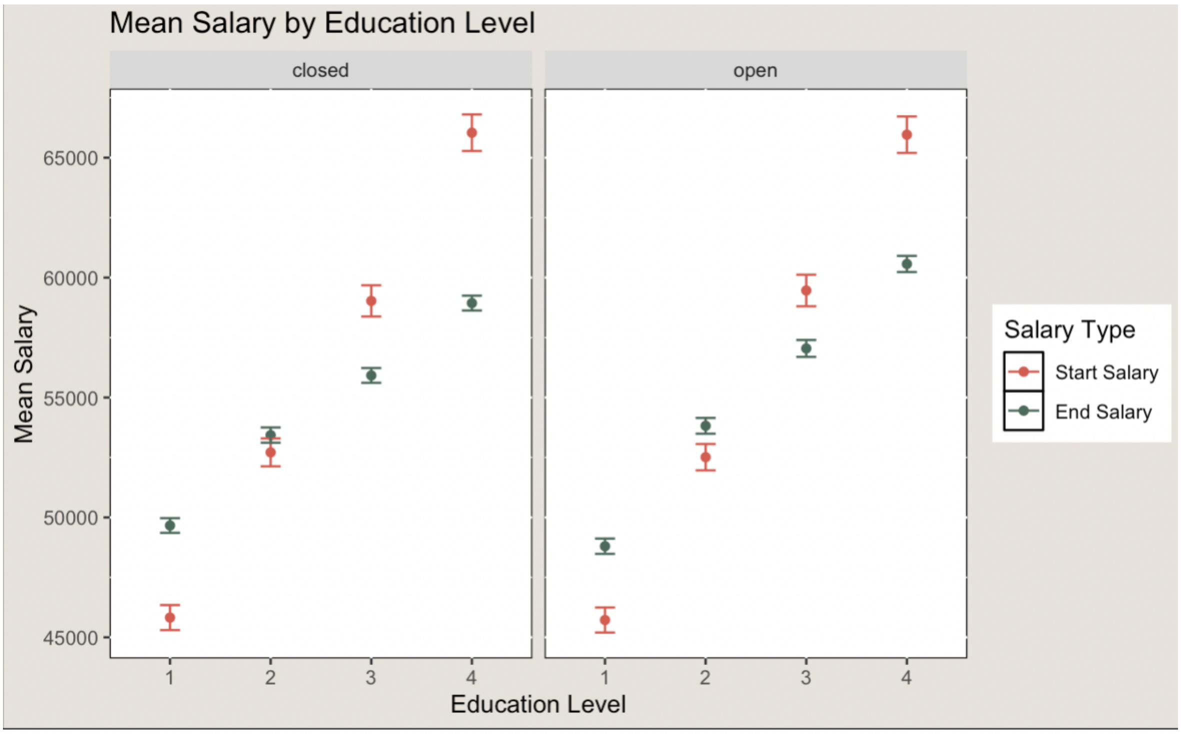

Figure 4 - Comparison of Mean Salary by Education Level Pre- and Post-Negotiation

3.3c Mixed Effects Model - Salary Difference (LD)

In order to investigate how agents who were under- or overpaid compared to their expected salary were affected by pay-transparency, a binary variable referring to whether the agent was overpaid in relation to their expected salary or not was first created. Simply, if their pre-negotiation salary was higher than their expected salary, they would be considered "overpaid". This resulted in roughly a 50/50 split[3]. To investigate how the salaries of these two new groups (overpaid and underpaid) were affected by the negotiation, a mixed effects model was employed. The dependent variable was the salary difference pre- and post-negotiation i.e. the raise/decrease the agent experienced. As independent variables an interaction effect between the 'overpaid' and 'condition' variables was used. Lastly, as a random effect, the Run-ID of the loop was used.
The syntax of this model can be seen below:

Salary Difference ~ Overpaidness[4] * Condition[5] + (1 | Run ID)

The results of this model can be seen in the table below.

---

[3] For a full summary of the number of agents in each group, as well as seeing mean salary pre- and post-negotiation for these two groups, see tables 3a and 3b in appendix.
[4] As mentioned, binary variables where 1 = a pre-negotiation salary > calculated expected salary and 0 = pre-negotiation salary < calculated expected salary.
[5] Transparent or Non-Transparent

Salary Difference

Beta Coefficients, Standard Errors, and P-values

Condition	Beta	Std. Error	P-value[1]
Non-Transparent & Underpaid	5183.7	241.2	< 0.001***
Non-Transparent & Overpaid	-9324.1	353.2	< 0.001***
Transparent & Underpaid	452.3	342.1	> 0.05
Transparent & Overpaid	-155.9	497.0	> 0.05

[1] P-values indicate significance levels

*Table 3 - Model outputs of salary differences**

**Please note that all beta-values are in relation to the intercept; condition non-transparent & underpaid*

Interpreting these numbers, the workers who were underpaid in the non-transparent condition saw a significantly higher salary difference compared to agents who were underpaid in the non-transparent condition. However, these differences were not significantly different compared to the transparent condition. The results are visualized for clarity in figure 5.

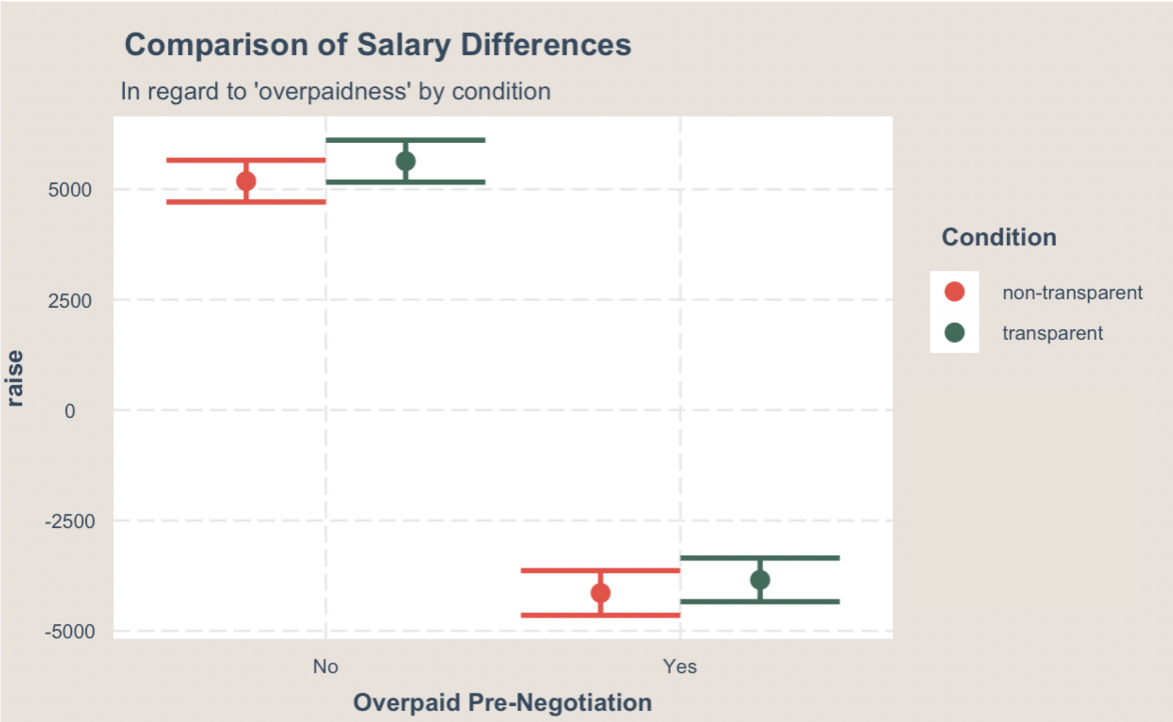

Figure 5 - Comparison of Salary Differences Based on Relative Earnings ('overpaidness') of Agent Per Condition

4. Discussion & Limitations

4.1 Discussion of Results (VM)

Ultimately, the results of these models are not fully in accordance with the expected results.

We hypothesized that agents in the transparent condition would see a lower mean average compared to the non-transparent condition; oppositely we saw the reverse effect. Both conditions increased in mean salary, but the transparent condition was even more so. Furthermore, although the expected result of seeing a tightening of the salary spread was seen in the transparent condition, i.e. relative low-earners saw an increase in their salaries, while relative high-earners experienced a decrease, this trend was also seen in the non-transparent condition (figure 5). This is also consistent with our results in the second model (§3.3b), showing that the less educated the agent was (and therefore, on average, a lower earner), the more the agents benefited from negotiations. The two lowest educated groups out of the four saw an increase in their wages (while agents with a high school diploma saw a higher increase than agents with a bachelor's degree), while agents with a master's degree saw almost no difference and agents with a PhD saw a decrease in their wages. As stated, these results were not expected to be present for both conditions.

For one, only a trend towards more equal payment was expected across the workplace in the transparent workplace condition, as low-earning agents were anticipated to have stronger arguments in the negotiation rooms if they knew they were being underpaid compared to their expected salary or in comparison with their co-workers. Therefore, it is strange seeing this phenomenon occur in both conditions, which may be a symptom of a flaw in the GABM used. Furthermore, a large number of agents (1720 out of 4657 to be exact) experienced a decrease in salary post-negotiation. While the salaries used in this paper were created relatively arbitrarily - which is not otherwise a problem for other aspects of the analyses in the paper, as it has mostly been of interest to examine the relative differences between conditions and groups - these extreme values are not ideal when trying to create the most naturalistic environment possible. These values, in addition to the large number of salary decreases post-negotiation, could stem from the prompted personality of the employer agent, whose goal was to keep the salaries of the worker agents as low as possible, and was simply too efficient at their respective goal. One may then imagine a different outcome of salary distribution if the goal of the employer agent was, for example, instead to be as fair and just to the workers as possible. Although speculation, this would explain why salaries decreased so much for high-earning agents for both conditions, and a more pronounced difference between conditions may be expected in a potential 'just-and-fair prompted employer' experiment instead.

The prompting of the agents involved creating a ToT for the agents which defined the problem space within which the agents could explore. However, the problem space was relatively narrow resulting in these similar results across different simulations. To address this limitation, a more diverse set of agents could have been generated with the ToT technique to prompt the creation of various types of employers, environments, and personalities. This approach would have allowed for a broader exploration of possible scenarios and outcomes although it would be more costly. The ToT technique could also have been deployed for the agents in their negotiations to deploy various negotiation strategies but cost is an important factor to also consider here. In general, prompt engineering is a crucial factor when working with GABMs. It allows the model to include a high degree of variability and complexity. However, it requires a careful balance, as poorly crafted prompts can lead the model to produce inappropriate or unintended results, which may also have been the case here (Giray, 2023; Yao et al., 2023).

4.2 Comparison with Previous Research

4.2a Results in Context of Previous Literature (LD)

The results of our simulation showed that while the mean salaries of agents in both conditions were positively affected (see figure 3), the transparent condition had a significantly higher post-negotiation salary compared to agents in the non-transparent condition. These results were not what we had hypothesized based on the previous literature on the matter, such as the aforementioned (§1.1) model created by Cullen & Pakzad-Hurson (2021).

They predict that *"pay transparency reduces individual bargaining power of workers, leading to lower average wages"* (Cullen & Pakzad-Hurson, 2021) as explained in §1.1. Oppositely, our results suggest that the transparent condition benefitted from a *greater* bargaining power allowing them to negotiate to a higher salary. Furthermore, our results also show that pay transparency affects the education groups differently (post-negotiation). The results show an increase in average salaries in the two groups with the lowest level of education and a decrease in average salaries in the two groups with the highest level of education. Cullen & Pakzad-Hurson (2021) reported that pay transparency might lead to greater salary decreases for employees with a longer education (4-year college degree) compared to employees with a shorter education, which is consistent with our results (§3.3b).

Overall, when it came to the predicted mean salaries for workers, our results differed from those expected by the model used by Cullen & Pakzad-Hurson (2021). When it came to education levels, however, our results were in accordance with their expectations.
However, this aforementioned paper includes many things in their mathematical models that are not included in our GABM. These include a maximum wage set by the firm, an "outside option" (the payoff they receive in case of unemployment, the probability of renegotiation based on the salaries of their co-workers, and other parameters. Potentially, further developing the complexity of this GABM used in this paper to include such parameters might yield a result more consistent with their results regarding pay transparency's impact on mean salary. Furthermore, the type of pay transparency used in this paper was also anonymous, meaning that agents only saw the mean salary of all co-workers and not each individual co-worker's salary, and results might therefore also have looked different if agents could see this information instead.

4.2b Results in Relation to EU Directive Objectives (VM)

One of the main goals of the pay transparency directive created by the EU is to ensure more pay equity based on the principle with a focus on the value of the work. Essentially, ensuring equal pay for work of equal value. However, with the trend of agents with lower education levels seeing an increase in their salaries and agents of higher education levels seeing a decrease, does on a surface level create a more "equal" and less spread out salary difference, this might lead to the opposite effect of the directive's intended goal. Assuming agents of higher education level provide higher-value work, these results instead suggest equal pay for work of unequal value. Nevertheless, it's hard to draw these conclusions from the results in this paper, as these results occurred for education levels in both conditions.

4.3c Cultural Aspect (SM)

Lastly, cultural aspects of negotiation (norms, behaviors, and beliefs) might also be relevant to include and consider in relation to pay transparency policies. This could be important because cultural negotiation patterns can help provide a better understanding of human behavior, as well as potentially helping make the policy approaches more diverse, adapted, and efficient. Gelfand & Brett, (2004) argue that communication processes are affected by cultural norms, goals, and behaviors. They argue, among other things, that high-context- and low-context communication as well as competitive- and cooperative goals can impact negotiation behavior (Gelfand & Brett, 2004, chapter 7 p. 161-163). For example, they argue that people from individualistic cultures tend to focus more on personal goals over social obligation, whereas people from collectivistic cultures tend to focus more on the social aspect (Gelfand & Brett, 2004, chapter 7 p. 159). This might then be reflected in their negotiation strategy, because people from individualistic cultures might prioritize individual gains over social cohesion, potentially leading to a more competitive negotiation behavior, whereas people from collectivistic cultures might prioritize social cohesion more, which could lead to a more collaborative negotiation behavior. Considering this, one may assume it might be more challenging for people from collectivistic cultures having to request salary information by approaching the employer themselves, compared to pay transparency policies where the information is shared without one having to request it. Nevertheless, these cultural differences are important to keep in mind when reflecting on the use of transparency laws in workplaces, indicating that such directives may not necessarily be a "one-size-fits-all" solution.

However, it is important to note that the cultures of negotiation are dynamic and complex and that they cannot be generalized, because identified patterns might not be representative of all cultures, but it might help give an insight into overall patterns that could be relevant to consider when creating pay transparent policies. Lastly, for a future GABM like the one used in this paper, these cultural differences could be interesting to investigate or at least include in the model to make it more naturalistic and diverse.

4.3 Methodological Limitations (VM)

This section will explore various aspects of the usage of GABMs which need to be considered. One of the main skepticisms regarding the use of large-language models in this fashion is the question of whether the behavior of an LLM accurately reflects the behavior of a human. If this is not the case, then critics might argue that the results of simulations like the one used in this paper cannot be used to infer anything about complex human social behavior. Instead, they will only reflect the behavior of LLMs, and the primary interest, for the most part, lies in studying *human* social dynamics, not artificial ones (Binz & Schulz, 2023).

While only very few would claim that large-language models' behavior equals human behavior, it is possible that the information gained from using GABMs can help us understand complex, dynamic situations. For example, the use of regular ABMs have been used in several instances to describe complex behavior, e.g. the COVID-19 papers discussed earlier (Cuevas, 2020, p. 19; Kerr et al., 2021; Kumaresan et al., 2023). Some then argue that GABMs are simply an evolution of the standard ABM. In his paper, Junprung (2023) describes "the final frontier for simulation is the accurate representation of complex, real-world social systems.", and explains that, while powerful, ABMs still are "unable to faithfully capture the full complexity of human-driver behavior". He posits that the emergence of powerful large-language models may be the solution to this problem of ABMs, "enabling researchers to explore human-driven interactions in previously unimaginable ways". (Junprung, 2023)While this is a very optimistic take on the use of GABMs, there are still several limitations of using this method compared to standard ABMs.

Namely, the resources required for running GABMs make the process long and potentially expensive. In total, running just 500 simulations took approximately 15 hours and cost around 20 USD - this is excluding the cost and time of trial runs. While this process could have been made faster by using the ChatGPT-4o model, as well as potentially making the behavior more natural (as it is trained on more data), the cost of doing so would have increased approximately tenfold. In essence, the price and the time it takes to run GABMs make the scalability quite difficult or, at the very least, heavily impractical. As we are still in the dawn of the boom of large-language models, it is hard to imagine these problems not becoming smaller still as the underlying technology advances. However, for now, scalability still presents a challenge in the use of GABMs.

Additionally, another problem presented by GABMs is that the output of these are long pages of text which can be harder to draw any conclusions from or do statistical analyses on; at the very least it requires some creative thinking (or utilization of NLP techniques) compared to analyzing the pure number output of standard ABMs. In our case, it required all our data to go through our logging-LLM which acted as a sort of filter, so as to avoid reading through all 5000 negotiation rounds and manually transcribing the post-negotiation salary. However, this comes with the danger of giving up some percentage of control and trusting the capabilities of this LLM. Although we did not act wholly in blind faith, and instead ran several trial runs, in which we gradually adjusted the prompt to minimize mistakes we saw, as well as setting the temperature to the lowest possible value - we still did not check through for all 5000 negotiation rounds whether the logging of the data was correct, which means a relinquishing of control is the price required to gather this large amount of data.
Furthermore, despite all these controls and checks for mistakes, large-language models are far from perfect and still regularly make mistakes and even 'hallucinate', which will be discussed in the following section.

4.4 Reliability and Validity (SM)

One key limitation of using LLMs for simulation, which this study also encountered, is their tendency to occasionally "hallucinate". A hallucination in terms of LLMs is when the model is generating context that branches off from the user input. The hallucinations of LLMs can be distinguished into three different categories: input-conflicting, context-conflicting, and fact-conflicting hallucinations. The input-conflicting hallucinations are hallucinations where the model deviates from the user input message and the response is an answer to something else. Context conflict is when the model suddenly deviates from the context and starts generating content not relevant to the context. The fact conflict is when the agent generates content that is not factual and conflicts with either real-world domain knowledge or database knowledge (Zhang et al., 2023). One of the methods used to mitigate the issues of hallucinations involved following some of the current best practices for prompt engineering (Giray, 2023). Some of the ways the study used prompt engineering included being very specific with the task at hand for the agents such as *"Please be concise and use very specific numbers when discussing salary. Do not accept offers that are not in concrete numbers."* and very explicitly instructing them in the scenario they were placed in. Despite this prompt engineering, agents still hallucinated and especially context-conflicting and fact-conflicting hallucinations were prominent. Some of the context-conflicting hallucinations encountered included agents suddenly shifting personas, forgetting about the setting, or not understanding the assignment. A lot of the fact-conflicting hallucinations included the agents forgetting their expected salaries, educational level, and being untruthfully about their effort level among others. However, some of the fact-conflicting hallucinations are scenarios that could also have happened in real-world scenarios such as employees being untruthful about their effort level and responsibilities which the model captures by chance. Another way to mitigate these issues could have been achieved by using a model with more parameters, such as Llama 3 or ChatGPT 4, which provide better performance (Zhao et al., 2023).

One potential issue of using GABMs lies in their reproducibility. Different models are pre-trained on different datasets and as of this, they are also generating content which may vary. The models also contain different biases due to the nature of the datasets they are pre-trained on which might shift the results in a way not intended (Hansen et al., 2024). Additionally, OpenAI have fine-tuned their model to behave very politely and to not break out of character. Examples of "jailbreaking" the model have been seen, but as of lately have been highly restricted by OpenAI, ensuring this is almost impossible. For the case of this study, it would have been optimal if it was possible to prompt the model without any of these limitations, as humans vary in behavior (Xie et al., 2023). A way to mitigate some of these issues could have been done by fine-tuning the model for the specific purpose. Fine-tuning is a method that involves feeding the model with targeted data and training it to act in a certain way (Howard & Ruder, 2018). This however would require gathering or generating a substantial amount of data from salary negotiations, ensuring that the model provides consistent and reliable results tailored to the scenario.

4.5 The use of GABMs in Social Sciences (LD)

While GABMs may be a powerful tool for very specific scenarios, they exist in an odd place between the studies on the chaotic, sometimes hard-to-measure naturalistic human interactions and dynamics and the complex, but highly-controlled, easily-scalable world of standard ABMs. GABMs lie in between these worlds, providing a more human-like interaction between agents in the form of dialogue, but, almost by definition, being less human-like than humans themselves. Although not fully human-like, GABMs offer a more practical, scalable, and time-efficient solution for simulating hundreds of environments with varying parameters compared to conducting real-world experiments or analyzing multiple workplaces in person. Essentially, it is a jack-of-both-trades, offering nuanced and context-rich simulations, but a master of neither the naturalistic authenticity of human interactions nor the precise control of traditional ABMs.

In addition to having the strengths of scalability and control from ABMs, it is still less powerful in those two parameters than a standard ABM. Likewise for human studies, it contains dialogue very reminiscent of how humans may communicate with each other, although to such less of a degree that it is debatable whether it is possible to infer anything about human behavior in the first place. Indeed, although it exists in some place between these two methods, it is clear that GABMs are of course more alike ABMs than it is to naturalistic human behavior.
If GABMs and ABMs were to be compared using Hammond's (2015) description of the three key strengths (heterogeneity, spatial structure, adaptation and co-evolution) of ABMs, it is emergent that GABMs are quite similar in their structure and strengths, although to different degrees. Looking specifically at the GABM used in this paper, it does not utilize these strengths. For heterogeneity, although the agents in this paper had different education, effort, and responsibility levels, only a maximum of 36 personalities[6] can

[6] This number increases double-fold to 72 if you consider the inherent gender of each agents' name.

be created from these combinations. However, it is indeed very possible to create a more diverse pool of agents and give them more unique traits in a GABM. For example, one could imagine giving agents different scores in each of the so-called 'Big Five' personality traits to create a persona (Goldberg, 1990). This was considered for our project, however, it was ultimately decided against as we felt it would add an extra layer of complexity which, although may have been interesting to investigate, was not the main focus of this paper. Similarly, the idea of including gender in our analysis was also considered, but this too was an extra dimension that would not necessarily help us understand how a pay-transparent workplace affects wages, despite its interesting nature. Research done by Liang et al., (2021) shows that LLMs might have inherent social biases such as gender, and more research on the field is needed on how these inherent biases might have influenced the results of this study.

While the heterogeneity of agents in this paper may be relatively low, some packages like Concordia, which are created to help design simulations using GABMs, run each agent through a "memory-maker" process in which agents are given simulated memories created by a LLM based on the attributes assigned by the researcher (Vezhnevets et al., 2023). While this package was considered for the creation of the GABM in this paper, this was decided against as the scalability of such a model would have resulted in far, far fewer possible simulations as the resources and time taken to run such thorough personality creation for each would be substantially higher. Ultimately, GABMs have the capability to create wildly diverse and heterogeneous agent pools, but at the cost of resources and time for each extra layer of complexity.

Comparing Hammond's second defined strength of ABM to a GABM, spatial structure, this is also a strength shared by GABMs. As previously mentioned, GABMs have been used to simulate all sorts of environments from broad environments like a community/society (Park et al., 2023), to more focused environments like a hospital with patient and doctor agents (Li et al., 2024). This latter paper segues nicely into another strength of ABMs and GABMs alike, namely the adaptation and co-evolution of environment and agents. In their paper, Li et al., (2024) show how doctor agents improve their diagnosing capabilities as the simulation goes on; in other words, how agents adapt and evolve in their environment.

In conclusion, while GABMs may be very similar to ABMs in many ways, and share many of the same strengths, they also have their own unique weaknesses, like scalability, 'hallucinations', and uncertainties of the equivalence of human behavior.

The question then remains; is there a place for the generalist middle child in social sciences? While the answer to how much this method can teach us about the social and cultural dynamics of humans is unknown (although at present its capabilities may seem limited), it is clear that with the rapid advancements in LLMs, in addition to the ever-increasing computing power of modern technology allowing for faster and more complex ABMs, the landscape for the further use and development of GABMs is fertile ground for potential advancements in social science.

4.6 Future Directions (LD)

This paper sets up many potential avenues to explore in the future. The obvious one is doing a semi-replication of this paper with greater resources (like ChatGPT4/ChatGPT4o) and a more complex environment. Concretely, the GABM used in this paper was quite simple - agents took turns negotiating their salary with their boss individually, meaning the worker agents did not actually interact directly with each other. In a future experiment, instead of creating many one-to-one scenarios between the employer agent and worker agents, one could imagine simulating a "one-to-many" office environment in which workers could interact and potentially strategize with each other before going into the negotiation room. Cullen & Pakzad-Hurson (2021) argue that *"in situations [..] .such as under collective bargaining agreements [..] greater transparency has a muted impact on average wages".* Therefore, this dynamic could be interesting to investigate.

Furthermore, continuing in the vein of a similar experiment, but with a different focus, a more diverse, heterogeneous agent pool could be created, which could lend itself to analyses of how different personality traits impact the ability to successfully negotiate a higher salary. There is also the matter of demographic factors such as gender and investigating whether large-language models have any inherent biases related to this and whether this affects the negotiation of salary. Furthermore, is this potential difference less pronounced when changing the proportion of males to females in the workplace, and would the gender of the employer have any impact on this? Safe to say, there are many paths one can go down with just this relatively simple environment.

It could also be interesting to create and include different workplace environments with different workplace cultures in the simulation, and run the simulation with similar agents but in different environments - for example, comparing collaborative and competitive workplaces, or comparing workplaces with either an individualistic or collectivistic culture.

4.7 Ethical Considerations (SM)

When working with AI, ethics are an important point to consider. To ensure that the research was in compliance with current ethical standards of using AI, we have been following the five pillars as outlined by IBM, 2021. The five pillars outlined by IBM consist of; explainability, fairness, robustness, transparency, and privacy. These guidelines emphasize the importance of using AI to benefit society, avoid harm, and ensure that systems are designed to respect human rights and freedoms.

Explainability: The model used and outputs can be explained through how LLMs work and the prompt engineering used for the model to behave in a certain way.
Transparency: Transparency has been maintained by documenting our methodologies. This detailed run-through of methodologies allows for replication of the study.
Fairness: Potential biases inherent in the model have previously been discussed in this paper. The model made by OpenAI has already been made to adhere to their ethical guidelines ensuring the model is not misused.
Robustness: The model was created to only provide specific results. Though a few deviations from the outputs were observed, the model generally produced consistent outputs.

Privacy: We ensured to not provide the model with any personal data and hereby we also adhere to any privacy concerns that could emerge by using AI.

Adhering to these principles is of importance when working with AI systems. Ensuring proper use is necessary since AI systems are developed to benefit society and minimize potentially harmful applications.

5. Conclusion (LD)

In this paper, the effect of pay transparency on salary negotiations and salary distribution was investigated, inspired by a recent EU directive on salary transparency as a means to combat wage inequality. This was explored using generative agent-based modeling, a new method of investigating social dynamics using generative large-language models as agents simulated in two conditions: one workplace which followed the guidelines of the aforementioned EU directive and one that did not. Agents in this environment took turns negotiating salaries with an employer agent in a one-on-one fashion. Contrary to expectations, a narrowing of the distribution of salaries was seen post-negotiation for both groups, although this change was hypothesized to only occur in the transparent condition. This unexpected result indicates that the post-negotiation salary may have been influenced more by the efficiency of the employer agent more so than the condition itself. However, a higher mean post-negotiation salary was observed in the transparent condition compared to the non-transparent condition, suggesting that pay transparency may lead to more bargaining power in the negotiation.
While this paper to some degree demonstrates the viability of using GABMs for simulating social interactions and dynamics, several limitations of this methodology were also identified. This included hallucinations of LLMs, the financial cost of running extensive simulations, as well as the time required to run many simulations. Issues of ensuring validity and reproducibility were also identified and discussed. Furthermore, the relatively simple environment, the efficiency of the employer agent, as well as a mostly homogenous agent pool, likely contributed to the unexpected results, which demonstrates a need

for more fine-tuned prompt engineering and agent diversity in future studies. Despite all these challenges, it must be noted that LLMs are in their early stages, and the aforementioned issues of GABMs will assumedly also grow smaller still as the development of large-language models continues to march forward. Bringing more human-like interaction aspect in the form of dialogue, while still maintaining the flexibility and scalability of a standard ABM (albeit to a smaller degree), suggests that GABMs may then still prove a useful tool for some areas of social science research, even if its capabilities may at the current time seem limited.

6. Code Availability (VM)

For the sake of transparency and the promotion of open science, all code - the GABM, the data collected therefrom, and the analysis code - can be found in the repository below.
https://github.com/VilliamJ/soccult_GABM

7. Appendix (SM)

Pre-Negotiation

Overpaid	# of Agents	Mean Salary	Standard Deviation
No	2450	48455.36	13434.45
Yes	2207	62965.81	16338.62

Table 3a

_

Post-Negotiation

Overpaid	# of Agents	Mean Salary	Standard Deviation

No	2450	48455.36	13434.45
Yes	2207	62965.81	16338.62

Table 3b

7. References

Avdul, D. N., Martin, W. M., & Lopez, Y. P. (2023). Pay Transparency: Why it is Important to be Thoughtful and Strategic. *Compensation & Benefits Review*, *56*(2), 103–116. https://doi.org/10.1177/08863687231181454

Bamberger, P., & Belogolovsky, E. (2010). The Impact of Pay Secrecy on Individual Task Performance. *Personnel Psychology*, *63*(4), 965–996. https://doi.org/10.1111/j.1744-6570.2010.01194.x

Berger, T., & Troost, C. (2014). Agent-based Modelling of Climate Adaptation and Mitigation Options in Agriculture. *Journal of Agricultural Economics*, *65*(2), 323–348. https://doi.org/10.1111/1477-9552.12045

Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, *120*(6), e2218523120. https://doi.org/10.1073/pnas.2218523120

Blake, L., & Peter, W. (2008). Introduction to the Special Issue on Agent-Based Models for Economic Policy Advice. *Journal of Economics and Statistics (Jahrbuecher Fuer Nationaloekonomie Und Statistik)*, *228*(2–3), 141–148.

Brady, M., Sahrbacher, C., Kellermann, K., & Happe, K. (2012). An agent-based approach to modeling impacts of agricultural policy on land use, biodiversity and ecosystem services. *Landscape Ecology*, *27*(9), 1363–1381.

https://doi.org/10.1007/s10980-012-9787-3

Columbia University, M. S. of P. H. (2016, August 3). *Agent-Based Modeling*. Columbia University Mailman School of Public Health. https://www.publichealth.columbia.edu/research/population-health-methods/agent-based-modeling

Council of the European Union. (2023a). *Pay transparency in the EU*. https://www.consilium.europa.eu/en/policies/pay-transparency/

Council of the European Union. (2023b). *Why pay transparency can help reduce the EU's gender pay gap*. https://www.consilium.europa.eu/da/infographics/pay-transparency/

CrewAI. (n.d.). *joaomdmoura/crewAI: Framework for orchestrating role-playing, autonomous AI agents. By fostering collaborative intelligence, CrewAI empowers agents to work together seamlessly, tackling complex tasks.* Retrieved May 23, 2024, from https://github.com/joaomdmoura/crewAI/tree/main

Cuevas, E. (2020). An agent-based model to evaluate the COVID-19 transmission risks in facilities. *Computers in Biology and Medicine*, *121*, 103827. https://doi.org/10.1016/j.compbiomed.2020.103827

Cullen, Z. B., & Pakzad-Hurson, B. (2021). *Equilibrium Effects of Pay Transparency* (Working Paper 28903). National Bureau of Economic Research. https://doi.org/10.3386/w28903

Dawid, H., & Neugart, M. (2011). Agent-based Models for Economic Policy Design. *Eastern Economic Journal*, *37*(1), 44–50.

Directive—2023/970. (2023, May 17). EUR-Lex. https://eur-lex.europa.eu/eli/dir/2023/970/oj

European Comission. (2021). *The gender pay gap situation in the EU - European Commission*. European Commission. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/equal-pay/gender-pay-gap-situation-eu_en

European Union. (n.d.). *Typer af lovgivning | Den Europæiske Union*. Retrieved May 31, 2024, from https://european-union.europa.eu/institutions-law-budget/law/types-legislation_da

Gelfand, M. J., & Brett, J. M. (2004). *The Handbook of Negotiation and Culture*. Stanford University Press. https://books.google.dk/books?hl=da&lr=&id=xsOfOIzoZigC&oi=fnd&pg=PA158&dq=How+does+workplace+culture+influence+wage+negotiation+outcomes&ots=qixEVU3OMP&sig=j1umUezXe4qneCFN9xUzwszUz7U&redir_esc=y#v=onepage&q=wage&f=false

Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering*, *51*(12), 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Hammond, R. A. (2015). Considerations and Best Practices in Agent-Based Modeling to Inform Policy. In *Assessing the Use of Agent-Based Models for Tobacco Regulation*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK305917/

Hansen, L. H., Andersen, N., Gallifant, J., McCoy, L. G., Stone, J. K., Izath, N., Aguirre-Jerez, M., Bitterman, D. S., Gichoya, J., & Celi, L. A. (2024). *Seeds of Stereotypes: A Large-Scale Textual Analysis of Race and Gender Associations with Diseases in Online Sources* (arXiv:2405.05049). arXiv. https://doi.org/10.48550/arXiv.2405.05049

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Heckbert, S. (2011). Agent-based modelling of emissions trading for coastal landscapes in transition. *Journal of Land Use Science*, *6*(2–3), 137–150. https://doi.org/10.1080/1747423X.2011.558599

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., Wang, J., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., Ran, C., Xiao, L., Wu, C., & Schmidhuber, J. (2023). *MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework* (arXiv:2308.00352). arXiv. https://doi.org/10.48550/arXiv.2308.00352

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. In I. Gurevych & Y. Miyao (Eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 328–339). Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1031

Hua, W., Fan, L., Li, L., Mei, K., Ji, J., Ge, Y., Hemphill, L., & Zhang, Y. (2024). *War* *and Peace (WarAgent): Large Language Model-based Multi-Agent Simulation of World Wars* (arXiv:2311.17227). arXiv. https://doi.org/10.48550/arXiv.2311.17227

IBM. (2021, September 17). *What is AI Ethics? | IBM*. https://www.ibm.com/topics/ai-ethics

Junprung, E. (2023). *Exploring the Intersection of Large Language Models and Agent-Based Modeling via Prompt Engineering* (arXiv:2308.07411). arXiv. https://doi.org/10.48550/arXiv.2308.07411

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Núñez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., George, L., Jastrzębski, M., Izzo, A. S., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S. L., Bennette, C. S., … Klein, D. J. (2021). Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*, *17*(7), e1009149. https://doi.org/10.1371/journal.pcbi.1009149

Kumaresan, V., Balachandar, N., Poole, S. F., Myers, L. J., Varghese, P., Washington, V., Jia, Y., & Lee, V. S. (2023). Fitting and validation of an agent-based model for COVID-19 case forecasting in workplaces and universities. *PLOS ONE*, *18*(3), e0283517. https://doi.org/10.1371/journal.pone.0283517

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Li, J., Wang, S., Zhang, M., Li, W., Lai, Y., Kang, X., Ma, W., & Liu, Y. (2024). *Agent Hospital: A Simulacrum of Hospital with Evolvable Medical Agents* (arXiv:2405.02957). arXiv. https://doi.org/10.48550/arXiv.2405.02957

Liang, P. P., Wu, C., Morency, L.-P., & Salakhutdinov, R. (2021). *Towards Understanding and Mitigating Social Biases in Language Models* (arXiv:2106.13219). arXiv. https://doi.org/10.48550/arXiv.2106.13219

McKinney, W. & others. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56.

NYT: Størst lønspredning i den private sektor. (n.d.). Retrieved May 29, 2024, from https://www.dst.dk/da/Statistik/nyheder-analyser-publ/nyt/NytHtml?cid=27961

OpenAI. (2024, May 26). https://platform.openai.com

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). *Generative Agents: Interactive Simulacra of Human Behavior* (arXiv:2304.03442). arXiv. https://doi.org/10.48550/arXiv.2304.03442

Peeperkorn, M., Kouwenhoven, T., Brown, D., & Jordanous, A. (2024). *Is Temperature the Creativity Parameter of Large Language Models?* (arXiv:2405.00492). arXiv. https://doi.org/10.48550/arXiv.2405.00492

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Salgado, M., & Gilbert, N. (2013). *Agent Based Modelling* (pp. 247–265). https://doi.org/10.1007/978-94-6209-404-8_12

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist*, *26*(2), 145–159. https://doi.org/10.1037/h0030806

Tenhiälä, A., Chung, D., & Park, T.-Y. (2023). Procedural Pay Transparency, Motivational Climate, and Employee Outcomes. *Compensation & Benefits Review*, *56*.

https://doi.org/10.1177/08863687231216135

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762

Vezhnevets, A. S., Agapiou, J. P., Aharon, A., Ziv, R., Matyas, J., Duéñez-Guzmán, E. A., Cunningham, W. A., Osindero, S., Karmon, D., & Leibo, J. Z. (2023). *Generative agent-based modeling with actions grounded in physical, social, or digital space using Concordia* (arXiv:2312.03664). arXiv. https://doi.org/10.48550/arXiv.2312.03664

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, *18*(6), 186345. https://doi.org/10.1007/s11704-024-40231-1

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models* (arXiv:2201.11903). arXiv. https://doi.org/10.48550/arXiv.2201.11903

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Williams, R., Hosseinichimeh, N., Majumdar, A., & Ghaffarzadegan, N. (2023). *Epidemic Modeling with Generative Agents* (arXiv:2307.04986). arXiv. https://doi.org/10.48550/arXiv.2307.04986

Xiao, B., Yin, Z., & Shan, Z. (2023). *Simulating Public Administration Crisis: A Novel Generative Agent-Based Simulation System to Lower Technology Barriers in Social Science Research* (arXiv:2311.06957). arXiv. https://doi.org/10.48550/arXiv.2311.06957

Xie, Y., Yi, J., Shao, J., Curl, J., Lyu, L., Chen, Q., Xie, X., & Wu, F. (2023). Defending ChatGPT against jailbreak attack via self-reminders. *Nature Machine Intelligence*, *5*(12), 1486–1496. https://doi.org/10.1038/s42256-023-00765-8

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., & Narasimhan, K. (2023). *Tree of Thoughts: Deliberate Problem Solving with Large Language Models* (arXiv:2305.10601). arXiv. https://doi.org/10.48550/arXiv.2305.10601

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models* (arXiv:2309.01219). arXiv. https://doi.org/10.48550/arXiv.2309.01219

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). *A Survey of Large Language Models* (arXiv:2303.18223). arXiv. https://doi.org/10.48550/arXiv.2303.18223