Synthetic Duality: A Framework for Analyzing Generative Artifical Intelligence's Representation of Social Reality

Daniel Karell[a,*], Jeffrey Sachs[a], Ryan Barrett[a]

[a] *Yale University, New Haven, Connecticut, United States*

Abstract

The development of large language models (LLMs) has caused concern about their potential risks, including how their ability to generate texts plausibly written by a person could affect our shared perception of the social world. Yet, it remains unclear how best to assess and understand the models' influence on our understanding of social reality. Building on insights into how social worlds are represented within texts, we take initial steps towards developing a framework for analyzing natural language generation's (NLG) content, as well as its consequences for perceptions of social reality. We demonstrate our "synthetic duality" framework in two parts. First, we show that advanced LLMs can create, with minimal guidance, reasonable portrayals of actors and ascribe relational meaning to those actors – virtual social worlds within texts, or "Mondo-Breigers". Second, we examine how these synthetic documents with interior social worlds affect readers' view of social reality. We find that they change individuals' perceptions of the ideas and rhetoric of actors depicted in the documents, likely by updating individuals' expectations about the actors and their meanings. However, additional exploratory analyses suggest it is models' style, not their construction of "Mondo-Breigers", that might be affecting people's perceptions. We end by discussing how our study illustrates a methodological approach for using generative AI to conduct sociological research, as well as how NLG may unsettle structural notions of individuality. Namely, reimagining the duality of individuals and groups could help theorize growing homogeneity in an increasingly NLG-informed world.

Keywords: Duality, Socio-semantic networks, Mondo-Breiger, Large language models, Natural language generation

*Corresponding author

Email addresses: daniel.karell@yale (Daniel Karell), jeffrey.sachs@yale.edu (Jeffrey Sachs), ryan.barrett@yale.edu (Ryan Barrett)

1. Introduction

The development of large language models (LLMs), such as OpenAI's ChatGPT, has caused widespread concern about their potential risks and harms (Hinton et al., 2023; Metz, 2023; Roe and Perkins, 2023; Roose, 2023). Building LLMs can, for example, (mis)direct research effort away from natural language understanding and come with unsustainable environmental costs (Bender et al., 2021). However, much of people's apprehension is due to LLMs' generative ability, or their capacity to create original texts that could ostensibly have been written by a person. While this ability is not itself problematic, LLM users could near-effortlessly generate and circulate dis- and misinformation (Feuerriegel et al., 2023; Goldstein et al., 2024) or accurate statements that nonetheless amplify biases in training data (Acerbi and Stubbersfield, 2023; Motoki et al., 2023; Gillespie, 2024). In addition, the models' emphasis on the plausibility of output over other kinds of text quality means that LLMs can inadvertently interweave inaccurate information with information that "sounds right" (Mittelstadt et al., 2023; Millière, 2023). This "bullshit" (Frankfurt, 2005) could then simultaneously swamp the internet and be microtargeted at individuals without anyone being directly accountable for it (Schneier and Sanders, 2023; Hackenburg and Margetts, 2024). Moreover, attempts to distinguish human-generated, or "organic", content from LLM-generated, or "synthetic", content can be difficult for many people, potentially causing them to quit trying to critically evaluate synthetic content (Spitale et al., 2023).

Yet, despite the growing commentary and evidence that LLMs could affect our shared perception of the social world, it remains unclear how to assess and understand the models' influence on our understanding of social reality. After all, misleading and biased content that is cheaply produced and widely disseminated is not new (Woolley and Howard, 2017), and propaganda campaigns have long shaped collective perceptions of the world (Bernays, 1928; Ellul, 1973; Stanley, 2016). Furthermore, researchers have thus far focused on LLMs' promise for text-analysis tasks (Chae and Davidson, 2023; Gilardi et al., 2023; Le Mens et al., 2023; Lira et al., 2023; Webb et al., 2023; Ziems et al., 2023; see also Bonikowski et al., 2022) and supplementing social science databases (Argyle et al., 2023b; Bisbee et al., 2024). Increasing numbers of scholars have focused on the social consequences of LLMs, but their work typically examines either how synthetic text can shift individuals' views and attitudes (Kreps et al., 2022; Jakesch et al., 2023; Hackenburg and Margetts, 2024; Goldstein et al., 2024) or alter interactions among humans (Argyle et al., 2023a; McKee et al., 2023). Little of this research has studied the influence of generative AI (genAI) on individuals' perceptions of the surrounding world.

In this article, we take initial steps towards developing a framework for analyzing text produced by genAI and its consequences for shared understandings of social reality. Our framework builds on insights into how social worlds are represented within texts (Lee and Martin, 2018), which themselves draw on Ronald Breiger's (1974) now-classic conceptualization of duality. Specifically, we begin with the simple proposition that texts depicting relations between people, groups, or organizations (henceforth, "actors"), along with what these relations convey and what social roles and categories the actors represent, can influence readers' understanding of the "organic" social world. The magnitude and scope of the effect depends on how exactly the texts portray the actors and their meanings – in other words, the texts' social and semantic network rendering of the social world. Lee and Martin (2018) call these "virtual social worlds" within texts "Mondo-Breiger"s; we refer to them as in-text socio-semantic (ITSS) systems.

We further posit that because the latest LLMs perform well at named entity recognition, natural language inference, and other reasoning tasks (*e.g.*, Bang et al., 2023; Webb et al., 2023; Ye et al., 2023), they can generate, with minimal human guidance, texts containing reasonable and realistic ITSS systems (see also Kreps et al., 2022). Finally, it is these cogent, yet not necessarily "true", ITSS depictions of actors and their meaning-based relations that enable genAI output to shift readers' understanding of the social world. That is, because LLMs' ITSS systems are passable as "organic" but also different than what humans create, they can slightly alter readers' perception of the world, such as what they think about the actors' ideas and the social categories they represent.

We demonstrate our "synthetic duality" framework for analyzing genAI in two parts. In the first part, we used one of the most advanced LLMs available at the time of writing, ChatGPT 3.5-Turbo, to generate text documents that portray actors in social and discursive interactions. Specifically, we created plausible transcripts of a news talk show that would be found on a politically conservative cable news channel, such as Fox News Channel (FNC). While the generation involved minimal guidance from us, it was not a straightforward task because of ChatGPT's constraints, such as a limit on generated tokens per prompt, which had the secondary consequence of causing the model to "forget" what had been generated earlier. Therefore, explaining how to create realistic complex synthetic documents like show transcripts is an additional contribution of our article.[1]

After creating the synthetic transcripts, we compared them to organic transcripts from a news talk show on FNC, *Fox & Friends*, using an approach to analyze ITSS systems that we term "semantic graph edit distance". The results indicate that ChatGPT is adept at generating documents with realistic ITSS when receiving the most minimal guidance from humans. In a supplemental analysis, we asked human annotators

[1]Previous research has generated simpler documents, like news reports (Kreps et al., 2022). We build on some of these earlier insights for constructing documents, but have to innovate because of the more elaborate structure of talk show transcripts.

to assess the organic and synthetic transcripts. Many of the annotators reported that the synthetic transcripts (and organic ones) were actual transcripts of a show that was in fact broadcast, and a majority responded that the synthetic transcripts (and organic ones) were similar to a transcript of a news talk show broadcast on a conservative cable news channel. The ITSS comparison and annotators' responses together support our first finding: LLMs can render reasonable and realistic social worlds in generated text.

In the second part, we examined how the synthetic transcripts affected readers' perception of the social world, and how this effect was different than the effects of both organic transcripts and manipulated synthetic transcripts, or synthetic transcripts that ChatGPT did not "want" to produce as its default. This part involved randomly assigning human study participants to read one of three transcript types: the organic transcripts; the "real", or unmanipulated, synthetic transcripts; or the manipulated synthetic transcripts. Based on participants' responses to questions before and after reading the transcripts, we find that the real synthetic transcripts influenced how people perceived politically conservative actors and rhetoric, but not the other types of transcripts did not. Specifically, our results indicate that some people who read the real synthetic transcripts came to think that conservative arguments were more logical (regardless of whether they agreed with the arguments) and that conservative television programming was more convincing to viewers (regardless of whether they concurred with the programming).

After presenting our results, we discuss how the texts' characterization of actors and their meanings, or the *style* of the ITSS systems, may influence readers. This discussion suggests that development of our framework should consider how to incorporate dimensions of style along with actors and semantic similarity in analyses of ITSS systems. Our insights also urge future work on genAI's risks to not only focus on misleading and toxic content, but also *how* LLMs' presentation of content – a function of its design and learning (or, "alignment") – could unintentionally make actors, ideas, opinions, and arguments more acceptable to people who would have otherwise dismissed them.[2]

Finally, our article also offers broader methodological and theoretical contributions. Methodologically, it describes an approach that uses genAI content as an instrument for conducting sociological research, taking up recent calls to explore how generative AI can advance sociological scholarship (Jensen et al., 2022; Bail, 2023; Davidson, 2023). With regard to sociological theory, our study suggests that scholars consider how the growing use of genAI may unsettle duality's traditional notion of individuality. This notion conceives of an individual as (in part) an expression of the combination of the social groups to which they belong, which are a subset of all available social groups. In contrast, LLMs, as we discuss below, can produce

[2]Whether increased acceptance of previously rejected content and actors is "good" or "bad" is another question.

texts depicting individuals as expressions of the intersection of a previously unrealistically large subset of possible social groups, potentially resulting in portraits of people as more generic than they otherwise would be. Consequently, the next generation of scholars building on Breiger's work may have to re-imagine what duality – and its foundational insights from Simmel (1955) – can tell us about a growing homogeneity in an increasingly genAI-influenced world.

2. Duality and the socio-semantic systems of generative AI texts

Breiger's (1974) pathbreaking exploration of duality adhered closely to Simmel's (1955) original focus, the interplay of people and groups. In brief, individuals and groups co-constitute one another: people – individuals' individuality – are, in part, defined and influenced by the social groups they belong to, while at the same time (and on the "dual" side), groups are defined by their individual members. Over the last quarter century, however, this insight has been applied to symbolic and cultural relations (*e.g.*, Carley, 1994; Mohr, 1998; Rule et al., 2015; Light and Odden, 2017), as well as socio-semantic or socio-symbolic systems (*e.g.*, Roth and Cointet, 2010; Basov and Brennecke, 2017; Fushe et al., 2020). In the latter case, social actors and cultural elements form the dual sides. On one side, social connections are defined by actors' shared use of ideas, symbols, and other cultural elements; on the other, these elements gain meaning by how actors use them with, or connect them to, other elements (Basov et al., 2020). This socio-cultural framework is readily applicable to people interacting in the physical world and constructing meaning systems shared by other living people (*e.g.*, Basov, 2020; Karell and Freedman, 2020), but it can also be used to describe the co-constitutive nature of social relations and meanings *within* cultural products or forms (Lee and Martin, 2018). Cultural forms, such as books, contain within them social worlds – internal socio-semantic systems, or what Lee and Martin (2018) term "Mondo-Breiger" structures – which the creators of cultural forms construct when they relate actors, such as characters in their books, to one another by how they are discussed. That is, cultural forms' social worlds, or ITSS systems, are defined by how actors are connected, and these connections (or lack of connections) exist because of the similar (or dissimilar) meanings the creators ascribed to the actors.

With the emergence of digital text, large corpora, and computational tools and techniques, social scientists have increasingly used representations like ITSS systems to study the social and semantic relations within texts created by people (for a recent review, see Stuhler 2022). This interest is due, in part, to the reasonable assumption that people depict socio-cultural structures in (some) texts that they perceive in their ("physical") lives and which are meaningful to them. But, recent testing of genAI models indicates that humans may not be the only content creators with a predilection to rendering social worlds in certain kinds of text. Some LLMs may also be adept at inscribing within texts various kinds of social and semantic relations, including ITSS systems, which are the focus of this study.

In-text socio-semantic systems comprise nodes and edges, which are defined during the writing – or, in the case of LLMs, the generation – of the text. The first component, nodes, are actors, such as people or groups in novels, roles in plays, and speakers in transcripts. We expect that recent LLMs, such as ChatGPT 3.5-Turbo, would do well at identifying relevant actors and placing them in generated text. Ye et al. (2023) show that ChatGPT's 3.5 models have become far better than previous models at entity detection, a common NLP task requiring the recognition and description of entities, such as people and organizations – in other words, actors. Put differently, the latest LLM models have learned about many potential actors; they "know" what these actors are, what they are called, and how to recall them when prompted for a particular kind of actor.

The second ITSS component, edges, connect actors. These connections are based on actors "being discussed at similar points" (Lee and Martin, 2018, 22). Creators of text induce socio-semantic relations among actors when they use their actors to represent similar ideas, or endow them with associated meanings. These meanings, and their associations with one another, depend upon the words surrounding the actors in the text (see Arseniev-Koehler, 2022b). For example, if two characters in a novel are regularly discussed in the context of plotting harm against other characters, they would be connected through their mutual representation of "evil doing" or "villain".

As with generating actors, we expect that many advanced LLMs, after prompted to consider a specific actor, would typically place that actor in relevant and appropriate contexts. Our expectation is based on tests of ChatGPT models indicating that the models frequently successfully accomplish natural language inference tasks (Ye et al., 2023; see also Laurer et al., 2023) and deductive and inductive reasoning tasks (Bang et al., 2023). These abilities mean that if the model is prompted with, for example, "Hayley is a politically conservative pundit. What would Hayley say about the right to own guns?" or "What would a gun control advocate say about Hayley?", it would place the actor Hayley in a reasonable and realistic context of words.

Our two expectations together suggest that the newer ChatGPT models are capable of generating socio-semantic systems in texts. They should be able to recall particular actors, then infer reasonable contexts – surrounding words – for these actors. As a result, some actors should be connected to others based on what they mean within the world of the text, or the text's internal social worlds. We reformulate our expectations as our first hypothesis:

- **H1**. If ChatGPT and humans produce the same type of text document, ChatGPT, with minimal guidance, will generate ITSS systems which are comparable to the ITSS systems in the human-created texts.

Recent theoretical advancements in communication and shared meaning help explain why genAI texts with plausible ITSS systems could influence our collective understanding of the social world. Namely, when readers encounter these kinds of documents, they are seeing a set of relationships between actors. These connections not only portray interactions, but also reflection what the actors mean vis-à-vis one another and what they mean in the broader social context (*e.g.*, as representations of social roles and categories) (Fushe, 2022; see also Mische, 2011; Pachucki and Breiger, 2010; McLean, 2017). As a result, the genAI depictions of ITSS systems inform readers' understanding of how the actors in the text relate to one another, what their identities are (with regard to one another), and what the social roles, categories, and collective identities represented by the actors mean. The LLM's construction of socio-semantic networks becomes, to the reader of the text, new information about what to expect about particular actors, thereby potentially updating the reader's perception of the social world (Fushe, 2022; see also Karell et al., 2023). For example, generated transcripts of a television talk show with plausible ITSS systems may not only influence readers' impression of particular actors on the show and how they interact with one another, but also their expectations of actors on the show in general, what the abstract roles or categories of "hosts", "discussants", and "guests" on the show entail, and what general kinds of ideas these actors, roles, and categories tend to express.

While providing readers with a novel depiction of actors' interactions and meanings may shift their beliefs about the world, we are skeptical that reading one or a few genAI texts will in fact change individuals' existing values or ideology (Kreps et al., 2022; Argyle et al., 2023a; Spitale et al., 2023; but see Jakesch et al., 2023). Instead, we think it is more likely that encountering synthetic ITSS systems can influence readers' expectations about actors, including how they tend to act and interact, as well as the meanings, ideas, roles, and categories these actors represent and how others perceive the actors and their meanings (Fushe, 2022). In other words, individuals' exposure to new (synthetic) information about the social world through genAI ITSS systems will primarily manifest as a shift in perceptions of what actors are like and what they generally mean in society.

- **H2**. Reading ChatGPT-generated text with organic-like ITSS systems will change readers' impressions of the actors and actors' ideas which are depicted in the ITSS systems, as well as what the actors mean to the reader and to others.

3. Data

Our study draws on two kinds of data: (1) text documents and (2) responses offered by people participating in online crowdsourced tasks. The former are transcripts of cable news shows typically found on channels like FNC and CNN, including a talk show informed by current events. These transcripts can be one of three types. The first type comprises transcripts of episodes of three shows broadcast on FNC during 2020: *Fox & Friends*, a multi-guest news talk show; *Tucker Carlson Tonight*, a news entertainment show featuring a media personality (in this case, Tucker Carlson); and *Fox News @ Night*, a show that primarily reports major events and other news. For simplicity, we sometimes refer to transcripts of this type as "true" or, following Spitale et al. (2023), "organic". The second type are transcripts of a multi-guest news talk show generated by an LLM with minimal human guidance. Since these synthetic transcripts are effectively the default genAI output, we sometimes refer to them as "real" synthetic transcripts. Doing so helps distinguish them from the final type of transcript, manipulated synthetic transcripts. These are transcripts produced by the LLM under specific conditions that compelled it to generate transcripts with particular (non-default) characteristics.

3.1. Organic cable news talk show transcripts

We collected every available transcript of *Fox & Friends* ($n = 135$), *Tucker Carlson Tonight* ($n = 247$), and *Fox News @ Night* ($n = 244$) broadcast during 2020 from ProQuest. We also collected transcripts of other 2020 shows broadcast on FNC ($n = 2192$), CNN ($n = 10954$), and MSNBC ($n = 1901$). To obtain the transcripts, we queried ProQuest using the names of the show's regular hosts in the byline search field, then downloaded the text files. For details, see Lanning et al. (2021), who detail the procedure we followed.

Our analysis focuses on the *Fox & Friends* transcripts. The show's multi-guest talk show format is well suited for examining the construction and effects of ITSS systems. Namely, they provide observations of different numbers of actors interacting various numbers of times, thereby allowing use to measure and examine ITSS systems with a range of nodes and edges. In addition, they are short enough to generate well with available LLMs, unlike, say, novels.[3] We use the *Tucker Carlson Tonight* ($n = 247$) and *Fox News @ Night* transcripts as benchmarks when comparing the real synthetic transcripts to the *Fox & Friends* transcripts (*i.e.*, the test of H1), and we use the CNN and MSNBC transcripts, along with the FNC ones, to train an embedding model used to measure semantic similarities in our text documents (see Footnote 13).

[3] A possible alternative type of text are film scripts or plays. However, readers of these types of (generated) texts may have recognized them as fiction, and thus have been less likely to alter their perceptions of the portrayed actors and their meanings.

STEPIEN: Look, it's – it's – it's scary times in – in – in Democrat- controlled states and cities. You know, Vice President Biden's own party, Ayanna Pressley, member of the squad, talking about encouraging, quote, "unrest in cities." Democrat – Democrat mayors refusing federal support and help that the president's offered ... and – and the – the people in Vice President – Vice President Biden's party who are encouraging violence on the streets. You don't see many supporters of the president throwing bricks through windows or setting buildings on fire. DOOCY: You know, Bill, it's – it's become very clear that the two campaigns have different things they would like to talk about. You would like to talk about law and order and the violence in the street and how some of these Democrat-led cities, as the president said, are just spiraling into chaos...KILMEADE: Some stats show that 45 percent of Biden voters are very comfortable mailing in votes. They're going to mail it in. Sixty-six percent of Trump voters are going to show up in person. Which led this Bloomberg funded group to conclude that Trump may win on Election Day, but Joe Biden, as the ballots come in, and some allow ballots to arrive after Election Day in many to – too many battleground states things could switch.

Figure 1: Excerpt of *Fox & Friends* transcript (2020)

Figure 1 displays an excerpt from one of the *Fox & Friends* transcripts. The transcripts denote actors, or speakers, by their names (*e.g.*, "Kilmeade"). The text preceding the names usually indicates what was being said before an actor spoke while the text after the name indicates what the actor said.

3.2. Real synthetic transcripts

To generate the real synthetic transcripts, we used the OpenAI application program interface (API) to connect to ChatGPT 3.5-Turbo.[4] Once connected, we prompted ChatGPT to write the transcripts at its own discretion, which generally entailed asking it a set of questions, then using its answers to construct a transcript document.

3.2.1. Set-up

We began constructing the transcripts by obtaining information and text from ChatGPT. This entailed asking two foundational questions, which determined core features of each transcript, and two guiding questions, which provided information that we applied whenever generating any of the real synthetic transcripts. We asked the foundational questions at each initialization of a real synthetic transcript, and the responses could vary each time (although we observed that they did not vary much). These questions were:

- *Pretend you are creating a typical Fox News talk show, how many different participants should be on the show? Please provide a single numerical result. . .*

[4] We connected to the API with the `httr` `R` package (1.4.7).

- *Pretend you are creating a typical Fox News talk show with* **`nodes`** *participants. Give a list of topic-headlines that you would cover in a single show. Provide headlines that align with conservative viewpoints often found on Fox News...*

In the second question, **`nodes`** was replaced with the number of actors that ChatGPT provided in response to the first prompt. Then, after the first time we asked the foundational questions we prompted ChatGPT with the two guiding questions:

- *On average, how many times does each participant speak on a topic? Please provide a single numerical result.*

- *Provide a numerical range for the number of words a participant will use in a single speaking turn.*

As mentioned, we asked the second pair of questions only once, and used the responses to set fixed values for the creation of all the real synthetic transcripts. That is, while the first pair of questions prescribed for each real synthetic transcript how many actors would appear in it and what they would talk about, the latter pair of questions set, for all real synthetic transcript, the number of times each actor would speak and a range of values determining how long they would speak. Based on the answers to the second pair of questions, we constructed the transcripts such that each actor would speak an average of three times on each topic. For each speaking turn, participants would contribute to the conversation in w words or less, with w chosen randomly from between 10 and 30 words, with the possibility of also randomly contributing 50 and 100 words.

To keep track of participants, we assigned each actor a label – a person name plausibly used in the United States – which we randomly generated prior to creating each transcript.[5] We also chose at random one participant to be the host of the talk show, thereby assigning to each actor (now labeled with names) the role of either "host" or (other) "participant".

Once we completed the preceding steps, we prompted ChatGPT to construct discourse between the show host and the other participants. The discourse was constructed through an interactive and iterative procedure. Iteration was required to create a full length transcript because the ChatGPT model produced a maximum of 4096 tokens per prompt. Unfortunately, iteration also introduced the challenge of maintaining the model's memory of what had already been generated for the transcript. The ChatGPT version we used employs an ASSISTANT to "remember" what has been said and continue a conversation, and we lost this

[5]To select names, we used the `randomNames` (1.5-0.0) `R` package.

functionality by having to iterate. Therefore, whenever necessary, we had to condition ChatGPT with both a manually-imputed ASSISTANT prompt, which was text from the previous round of iteration, and a USER prompt, which identified the task we were asking the model to perform. Put differently, by feeding ChatGPT an ASSISTANT prompt with a USER prompt, we were able to first remind the model of what the previous participant on the talk show has said with the ASSISTANT prompt and then ask it to add to the conversation with the USER prompt.

In the following explanation of the remaining procedure for constructing real synthetic transcripts, we use boldface text, like **object_name**, to denote an object. For example, **previous** signifies the previous speaker in the conversation and **current** mark the current speaker in the conversation. Similarly, w, the word count, is denoted by **word_count**.

The remaining procedure to construct a transcript proceeded through these stages:

1. Start the talk show and/or introduce new topic.
2. Create talk show participants' viewpoints in an "exchange".
3. Repeat stages 1 and 2 for the number of topics.
4. End talk show.

3.2.2. Start the talk show, introduce new topic, and end talk show

As explained in Section 3.2.1, we set up the transcript creation procedure by asking ChatGPT for a number of actors and a set of topics (**topics**). The procedure then sequenced through all the topics of **topics** object beginning with the first.

On the first topic in **topics**, we gave ChatGPT the USER prompt: "Pretend you are the host of a Fox News talk show discussing the topic of **new_topic**. Your name is, **host**. Provide an introduction in less than 200 words using a perspective that aligns with conservative viewpoints often found on Fox News..."[6] Then, we iterated to create an exchange of actors' viewpoints, which we explain in *Creating talk show viewpoints*, below.

When transitioning to a new topic, ChatGPT was given the USER prompt: "Pretend you are the host of a Fox News talk show. Your name is, **host**. You and your guests have just finished discussing the topic of **old_topic**. Provide an introduction in less than 200 words to the new topic **new_topic** using a perspective that aligns with conservative viewpoints often found on Fox News..."

[6]Before the first instance of this prompt, we asked ChatGPT to tell us how long a host's introduction would be. We used the answer, 200 words, to design this prompt. We used the same technique to determine the length of the host's introduction of new topics and a conclusion. This is similar to the use of guiding questions, explained earlier.

Finally, when ending the transcript, we gave ChatGPT the USER prompt: "Pretend you are the host of a Fox News talk show. Your name is, **`host`**. You and your guests have just finished discussing the following topics: **`topics`**. Conclude your talk show in less than 200 words using a perspective that aligns with conservative viewpoints often found on Fox News..."

3.2.3. Creating talk show viewpoints

After the introduction of the first topic or a new topic, the actors participated in an "exchange". These exchanges were where the model portrayed actors as giving their viewpoints, and they formed the bulk of the transcripts. We compiled exchanges through the following steps.

1. We set **`previous`** to the host name and **`previous_text`** to the text created by the host when they introduced the first or new **`topic`** (see the preceding section).
2. We randomly selected the **`word_count`** for a new actor, or "participant", who will express a viewpoint.
3. We randomly selected a participant (other than the host). This participant was assigned the role of **`current`** speaker.
4. We prompted the **`current`** speaker with an ASSISTANT prompt: "The previous viewpoint on the talk show was from **`previous`**, a **`role(previous)`**[7]: **`previous_text`** on the topic of **`topic`**..." In this same step, we also prompted the **`current`** speaker with a USER prompt: "Pretend you are a guest named, **`current`**, on a Fox News talk show and respond to the previous viewpoint. Limit your response to less than **`word_count`** words. Add to the conversation with a new viewpoint but one that aligns with conservative viewpoints often found on Fox News..."
5. We set **`current`** speaker to **`previous`** actor.

The number of viewpoint expressions (*i.e.*, steps 1 through 5) in each exchange equaled three times the number of participants. The participants were selected randomly, and thus iterating from one to the number of participants three times resulted in each participant interacting an average of three times per topic. After each viewpoint expression, we selected from the set of **`available`** actors. The **`available`** set of actors was initialized as the entire set of participants, including the host. At the end of each viewpoint expression, however, we changed the **`current`** actor to the **`previous`** actor (step 5) and removed them from the set of **`available`** participants. This way, actors could not speak multiple times in a row.[8] When the number of

[7]The notation **`role(previous)`** is shorthand for checking whether the role of the previously speaking actor was either "host" or (other) "participant".

[8]We acknowledge that this choice perhaps overly simplifies some of the noise and dynamics of a real conversation, and future research could work towards better replicating greater conversational complexity.

HAYLEY: The growing bipartisan divide and the threats it poses to conservative values. As conservatives, we have always championed the principles of limited government, individual liberty, and free-market capitalism. However, in recent years, we have witnessed a dangerous shift towards radical progressivism, with the left pushing their agenda at an alarming pace ... ADAM: Hayley, you hit the nail on the head. The growing bipartisan divide is indeed a threat to conservative values. It's disheartening to see the left pushing their radical agenda without considering the consequences for our nation. We must stand strong in defense of limited government, individual liberty, and free-market capitalism. Now, let's delve deeper into the dangers of this divide. The left's relentless assault on our Second Amendment rights undermines our ability to protect ourselves and our families ... SAVANNAH: Adam, you're absolutely right. The Biden administration's failure to address the border crisis is putting our national security at risk. We need strong border security measures to protect American citizens and prevent illegal immigration. It's time to prioritize the safety and well-being of our own citizens.

Figure 2: Excerpt of a real synthetic transcript

participants in `available` reached one, `available` was restored with the set of all actors, including the host, and a second exchange began. The exchanges ended with the last topic in `topics`.

3.2.4. Completed transcripts

We generated 20 real synthetic transcripts. After creating the first few transcripts, it became clear that key features of the transcripts – the number of actors and the number of topics – were not varying much. In other words, ChatGPT had settled quickly on how many actors should appear each time a transcript was generated and what kinds of things they should talk about. Therefore, we were not gaining new variation in real synthetic transcripts' ITSS by generating more transcripts.[9] Figure 2 provides an excerpt of a real synthetic transcript. As with Figure 1, we can see the names of each speaker and the text capturing what they are depicted as saying.

3.3. Manipulated synthetic transcripts

We created the manipulated synthetic transcripts by conditioning ChatGPT to vary two dimensions: the number of actors, or speakers, shifting the number either to a high level and a low level, and the degree of discursive similarity among the actors' exchanges, also to high and low levels. These two manipulations produced four versions of the manipulated synthetic transcripts:

- many actors, high semantic similarity
- many actors, low semantic similarity

[9] This finding aligns with recent research showing that survey data generated with ChatGPT have less variation than organic survey data (Bisbee et al., 2024).

- few actors, high semantic similarity
- few actors, low semantic similarity

We determined a high number of actors in a transcript by first calculating the mean and standard deviation of the number of actors in the organic FNC transcripts. Then, we defined a high number of actors as the mean number of actors in these transcripts plus 1.5 times the standard deviation. We set the low number of actors in a transcript to two, since subtracting 1.5 times the standard deviation from the mean yielded a negative value. With respect to the number of actors, generating manipulated synthetic transcripts followed the same procedure as for the real synthetic transcript, with only the exchanges' iterations across actors becoming longer or shorter for high or low actor levels, respectively.

To manipulate the semantic similarity of actors' exchanges, we introduced new prompts to ChatGPT. For high similarity, we prompted ChatGPT with: "Pretend you are a guest named **`current`** on a Fox News talk show and respond to the previous viewpoint. Limit your response to less than **`word_count`** words. Add to the conversation with an extremely similar viewpoint. . . " Note that with this prompt, we conditioned ChatGPT to make the next contribution to the exchange a similar viewpoint.

For low similarity, we gave ChatGPT a different prompt: "Pretend you are a guest named, **`current`** on a Fox News talk show and respond to the previous viewpoint. Limit your response to less than **`word_count`** words. Add to the conversation with an extremely different viewpoint. . . " As is clear in the prompt's instructions, this prompt conditioned ChatGPT to contribute a different viewpoint from the previous one in the conversation. For both high and low similarity levels, we used the same ASSISTANT prompts as we did when creating the real synthetic transcripts.

3.4. People's responses to transcripts

Our second kind of data consists of people's responses to reading our transcripts. We gathered these responses from individuals participating in two online tasks, which we explain in the following section.[10] The first task, which we refer to as the "assessment task", had 252 participants. Their average age was about 39 years, most self-identified their race or ethnicity as "White", the majority voted in the 2020 presidential election, and all lived in the United States. The sample had roughly equal numbers of men and women. More identified themselves on the political left (141) than right (65); 46 identified as moderate. On average, the participants watched about an hour and a half of programming on a cable news channel per day. Table 1 further describes the sample.

[10] We implemented the tasks through CloudResearch, an online crowdsourcing platform commonly used for research.

The second task, the "response task", had 350 (different) participants. As we explain in the following section, these participants were randomly assigned to one of six groups; participants in each group read one type of transcript. In our analysis and discussion, we focus on the participants who read the organic *Fox & Friends* and real synthetic transcripts. Table 2 describes these two groups (which consisted of 61 and 60 participants, respectively). Their mean ages were about 44 and 42; they had roughly equal numbers of men and women; and everyone in these groups lived in the United States. As seen in the *p*-value column of Table 2, the groups do not differ overall in terms of several attributes, such as race or ethnicity, income, political ideology, the number of hours watching cable news channels, and participation in the 2020 presidential election.[11]

[11] A *p*-value greater than 0.05 indicates that we fail to reject the null hypothesis of no difference in means, suggesting that there is no statistically significant difference between the two groups.

Attribute	Statistics and counts	
Age	Mean (SD)	38.69 (11.37)
Gender	Female	117
	Male	130
	Non-binary / Third Gender	3
	Prefer not to say	2
Race/ethnicity	Asian	17
	Black	32
	Hispanic/Latino	16
	Native American	1
	Other	5
	White	181
Country	United States	252
Citizenship	Other	1
	United States	251
Employed	No	37
	Yes	215
Income	$0-30,000	35
	$31,000-60,000	47
	$61,000-90,000	74
	$91,000-120,000	56
	$120,000+	36
	I prefer not to answer	4
Political ideology	Extremely Conservative	13
	Conservative	32
	Slightly Conservative	20
	Moderate/Middle of the Road	46
	Slightly Liberal	32
	Liberal	66
	Extremely Liberal	43
Hours/day of cable news	Mean (SD)	1.33 (3.68)
Voted in last election	No	32
	Prefer not to say	5
	Yes	215

Table 1: Attributes of the participants in the assessment task

		True transcripts	Real synthetic transcript	p-value
Age	Mean	43.90	41.37	0.24
	SD	12.03	11.70	-
	N	61	60	-
Gender	Male	32	32	0.85
	Female	27	27	-
	Non-binary / Third Gender	2	1	-
Race/ethnicity	Asian	4	2	0.54
	Black	6	6	-
	Hispanic/Latino	2	5	-
	Native American	0	1	-
	Other	0	1	-
	White	47	44	-
Country	United States	61	60	0.93
Citizenship	United States	61	60	0.93
Employed	No	13	6	0.14
	Yes	48	54	-
Income	$0-30,000	14	6	0.18
	$31,000-60,000	17	19	-
	$61,000-90,000	8	16	-
	$91,000-120,000	10	7	-
	$120,000+	11	12	-
	I prefer not to answer	1	0	-
Political ideology	Extremely Conservative	4	5	0.76
	Conservative	6	9	-
	Slightly Conservative	6	7	-
	Moderate/Middle of the Road	13	10	-
	Slightly Liberal	6	10	-
	Liberal	16	15	-
	Extremely Liberal	9	5	-
Hours/day of cable news	Mean	1.60	1.28	0.51
	SD	2.46	2.84	-
Voted in last election	No	10	8	0.32
	Prefer not to say	2	0	-
	Yes	49	52	-

Table 2: Attributes of the participants in the response task

4. Analytical strategy

The analysis has two parts. First, we compare the organic *Fox & Friends* transcripts to the real synthetic transcripts using our semantic graph edit distance. This part also contains a supplemental test that compares the transcript using human assessments. In the second part, we examine how real synthetic transcripts could potentially shape our collective social reality.

4.1. Comparison of transcripts

We begin our analysis by testing H1 with a comparison of the real synthetic transcripts to the *Fox & Friends* transcripts. To do so, we first identified the ITSS systems in the texts. We calculated the number of actors by counting the unique speakers in the transcript (*e.g.*, Stepien, Doocy, and Kilmeade in Figure 1). We encoded semantic connections by measuring the similarity in actors' meaning. That is, actors were connected to one another if they were associated with similar meanings (in the text), akin to how characters in a novel could be considered alike if they evoke similar collective ideas, narratives, morals, or other meanings.

To measure the similarity in actors' meaning, we used an embedding approach. This approach, generally speaking, draws on learned models that quantitatively represent the semantic information among words in corpus. During leaning, an algorithm uses information about each word and its context – the surrounding words at each occurrence of the word – to discern the linguistic structure of the corpus. The linguistic structure is rendered as a multi-dimensional space, and each word is located in this space based on its semantic use in the corpus (Rodman, 2020; Stoltz and Taylor, 2021; Arseniev-Koehler, 2022a). The words' locations, or embeddings, are vectors with the same dimensions as the embedding space. We can compare these vectors to make inferences about semantic similarities among words: words with vectors that place them near one another are understood as having similar meanings, while more distant words have less similar meanings (*e.g.*, Nelson, 2021; Karell and Sachs, 2023).

We conducted the measurement step using *à la carte* (ALC) embeddings (Arora et al., 2018; Khodak et al., 2018; Rodriguez et al., 2023). The ALC technique does not embed a word of interest, but rather its context words – the words used alongside the target word. These context words' vectors are then averaged, creating one context vector for each occurrence of the target word, and then multiplied by a transformation matrix to downweight common words that might often appear across context windows (Khodak et al., 2018).[12] Since we are interested in the meaning of actors, we treated actors' names as the target words, thus computing

[12]These common words could include words associated with the spontaneity of oral discourse in American English, such as "uh" and "you know".

ALC vectors each time the transcript depicted an actor as speaking or being referred to by another actor. Our interest in actors is also why we used the ALC technique: it obtains high quality embeddings even when target words are relatively rare (Khodak et al., 2018), which is the case when transcripts capture a guest who speaks only a few times. After computing each actors' ALC vectors, we calculated the vectors' cosine similarity with each of the other actors' ALC vectors, then determined the mean cosine similarity. Higher similarity values indicate that two actors have more similar meanings in the text.[13] Finally, we created adjacency matrices for each organic and real synthetic transcript, in which the actors were arrayed across rows and columns, and their intersections contained their mean cosine similarity value. These text-level matrices captured each text's ITSS system.

Having identified each text's ITSS system, we then compare the two types of transcripts using our semantic graph edit distance approach. This approach entails counting which changes must be made to a comparison network, or graph, to make it match a baseline network. In usual graph edit distance applications, the core change options are inserting or deleting nodes and/or edges.[14] However, because Mondo-Breigers capture not only structure – nodes and edges – but also semantic information in their edges, we use variations of the typical add-or-delete edge count. Specifically, we calculate whether the comparison network's edges must increase or decrease their weight – the mean cosine similarity values – to become like the baseline network. If edge increases are necessary, this conveys some semantic insight: the text must endow its actors with more similar meanings. If edge decreases are required, the text must depict its actors with less similar meanings. Our assessment of nodes is the same as in usual applications: a count of how many nodes must be inserted or deleted.

The graph edit distance framework is descriptive; it simply conveys how one network must change to be like another network. Therefore, to evaluate whether our real synthetic transcripts' ITSS systems are notably like the organic ones, we use three benchmarks. First, we compare a randomly sampled half of the *Fox & Friends* to the remaining half.[15] The number of required node changes, edge increases, and edge decreases

[13] When implementing the ALC technique, we first trained a local embedding space and transformation matrix using a corpus comprising preprocessed transcripts from all shows broadcast on three major United States cable news channels – CNN, MSNBC, and FNC – during 2020 (N = 15673) using the `text2vec` package (0.6.3) for `R`. To compute the transformation matrix, we followed the procedure described by Rodriguez et al. (2023). Then, we preprocessed each *Fox & Friends* and real synthetic transcript by removing symbols, numbers, and other non-alphabetic characters, as well as common English stopwords and words with less than two characters, and tokenizing the text. To conduct this step, we used the `quanteda` package (3.3.1) for `R`. Next, we obtained the ALC vectors based a context window of six words; we used the `contText R` package (1.4.3). Finally, we calculated vectors' cosine similarities using `text2vec` (0.6.3). In a supplemental analysis, we computed speakers' ALC vectors based only on the words that followed the appearance of each actor's name – that is, only the words they were portrayed as saying. The results of the similarity calculations were consistent with those obtained when using the common context window of the surrounding six words.

[14] Other operations are possible, such as node substitution and edge splitting.

[15] In this and other comparisons, we compute the semantic graph edit distance between every network in the comparison set and every network in the baseline set.

should be low for any given pair of networks, and few comparison networks should require changes to be like those in the baseline set. The next two benchmark consists of ITSS systems from the *Tucker Carlson Tonight* and *Fox News @ Night* transcripts. Since these ITSS systems come from different types of texts – transcripts of news entertainment and news reporting shows – they should require more node changes, edge increases, and edge decreases for any given pair of networks, and more comparison networks should require changes to be like the baseline *Fox & Friends* networks.

We additionally compare the transcripts using people's assessments of the documents. To collect these assessments, we recruited individuals into an online task (see Section 3.4 and Table 1), explaining to the potential participants that the task entailed reading "cable news talk show transcripts".[16] Once participants joined the task and passed a test to ensure they were human, they read a randomly selected *Fox & Friends* transcript and a randomly selected real synthetic transcript. After reading each transcript, participants answered two questions about the plausibility of the transcript. The first question was:

> *Do you think this is an actual transcript from a news talk show that was in fact broadcast on a cable news channel?* (Question 1)

The possible responses were:

- Definitely not an actual transcript
- Probably not an actual transcript
- I am not sure if it is an actual transcript or not
- Probably an actual transcript
- Definitely an actual transcript

The second question was:

> *How similar is this transcript to what you imagine a transcript of a news talk show broadcast on a conservative cable news channel would be?* (Question 2)

Participants could respond with:

- Not similar at all

[16] We further explained that the transcripts depict "a news talk show produced and broadcast on an American cable news channel with a politically conservative lean. The transcript may be an actual transcript or one created independently to be like an actual transcript. The transcripts of actual show broadcasts were generated automatically with the same software that produces closed captioning."

- Not very similar
- I am not sure how similar it is
- Somewhat similar
- Very similar

At the end of the task, participants answered several questions about their demographics, political ideology, and media consumption.

4.2. Effects of synthetic transcripts

The second part of our analysis examines how genAI cultural forms might shape our social reality. This part consists of a small experiment, which we consider akin to a pilot. Our aim is to offer initial evidence and facilitate the development of more nuanced and expansive experiments. For this part, we recruited another set of participants to a second online task that collected their responses to reading a type of transcript (see Section 3.4 and Table 2). We explained that this task involved reading a transcript of "a cable news talk show" with a "politically conservative lean" that may or may not be an actual transcript generated by closed captioning software (see Footnote 16). Participants that opted into the task and passed a test to verify they were human, then answered two questions about how they perceived conservative actors, their ideas, and their programming. Specifically, we asked:

> *How logical do you think today's politically conservative viewpoints are, as you understand them? By logical, we mean whether they make sense and have clear reasoning, whether or not you agree with them.* (Question 3)

The response options were:

- Not logical at all
- Somewhat logical
- Very logical

> *A conservative cable news channel, Fox News, is the most-watched cable news channel in the United States. In addition, some academic research has shown that watching Fox News can affect its viewers' political attitudes and voting behavior. How convincing do you think the programming on Fox News is for a hypothetical average viewer of cable news? That is, do you think it is likely to have an impact on an average viewer?* (Question 4)

The possible responses were:

- There will be no impact
- There is will be little impact
- There will be a major impact

It is important to note that neither Question 3 nor 4 ask about participants' political beliefs, attitudes, or ideology. Rather, the questions probe how they perceive conservative actors and their rhetoric in the media. Namely, do they see it as logical and convincing, whether or not they agree with it?

We next randomly selected participants to read one of the transcript types: a *Fox & Friends* (organic) transcript; a real synthetic transcript; a synthetic transcript with high node and edge levels; a synthetic transcript with high node and low edge levels; a synthetic transcript with low node and high edge levels; or a synthetic transcript with low node and edge levels.[17] We use the manipulated transcripts as conditions for two reasons. First, to understand whether any detected effect is due to a basic difference between an organic transcript and any kind of a synthetic one. Second, if organic and synthetic transcripts do exhibit different effects, we hope to gain insight into what about synthetic transcripts might drive an effect. In other words, if we are correct that the ITSS properties of "naturally" generated synthetic texts matter, then the real synthetic transcript should affect participants differently than the manipulated transcripts.

After reading the randomly assigned transcript, participants answered Questions 3 and 4 again,[18] allowing for a within-subjects estimate of an average treatment effect.[19] Finally, participants answered questions about their demographics, political ideology, and media consumption. We use a paired sample t-test to test for the effect of reading a given transcript.

5. Results

We present our results in two parts. First, we show evidence that the LLM "naturally" generates synthetic transcripts that are similar to the *Fox & Friends* transcripts in terms of ITSS properties. This finding is further supported by the results of people's assessments. Then, we show evidence that the real synthetic

[17] The transcript document participants read was itself randomly sampled from each set of transcript types. Before selecting the documents, we refined the sets based on word length: since the mean number of words in *Fox & Friends* transcripts is 1620, we only sampled from the subset of documents which had at least 1000 words and truncated all selected documents at 2000 words (if they reached that number).

[18] After reading the transcript but before answering Questions 3 and 4, participants also responded to an attention check. The check asked how many unique speakers there were in the transcript.

[19] Comparing participants' post-treatment responses to their pre-treatment responses helps to avoid interpersonal incomparability biasing the results.

transcripts can influence how people perceive actors and their social meanings in ways that the true transcripts and manipulated transcripts do not.

5.1. How actual and synthetic transcripts compare

Figure 3 presents the test of H1, a comparison of the real synthetic transcripts to the *Fox & Friends* transcripts using the semantic graph edit distance approach. In the top row, we see that a random sample of half of the organic news talk show transcript does not require many changes to match the remaining organic news talk show transcripts. For any given pair of ITSS systems, few node insertions or deletions and few edge increases or decreases. Furthermore, the number of networks that do need changes are few in number. Taken together, these results indicate that the ITSS systems in the randomly sampled organic news talk show transcripts are similar to the ITSS systems in the other organic news talk show transcripts – exactly what we would expect.

The vertical dashed lines in Figure 3 mark the means of the distributions of required node and edge changes when comparing the two sets of organic news talk show transcripts. If the ITSS systems in the real synthetic news talk show transcripts are similar to those in the organic ones, then the distributions of the changes necessary to make the former like the latter should fall closely around the means marked by the dashed lines. This is what we find (see the second row of Figure 3). The number of necessary node changes and edge decreases are similar – both within and across texts – as the random sample of organic news talk show transcripts; the number of necessary edge increases is not as similar, but not drastically different.

To better understand the degree of similarity between organic and synthetic news talk show transcripts, we conduct two more comparison. In the bottom two rows of Figure 3, we see that the organic news entertainment and news reporting shows' ITSS systems would require more node changes and edge decreases (both within and across texts) than the synthetic transcripts to be like the baseline talk show transcripts. They would need relatively fewer node increases, but, as a whole, these benchmarks indicate that the "virtual social worlds" in the organic and synthetic talk show texts are more alike than the "virtual social worlds" in the organic transcripts. This evidence supports H1.

We supplement the results of the ITSS comparison with people's assessments of the similarity between the organic and synthetic talk show transcripts. Figure 4 shows the responses to Question 1. We see that the most common response is that both types of transcripts are "probably" actual transcripts. In addition, about 46% of task participants said that that the synthetic transcript was either "probably" or "definitely" a true transcript. While we would obtain more compelling evidence if a large majority responded this way, we

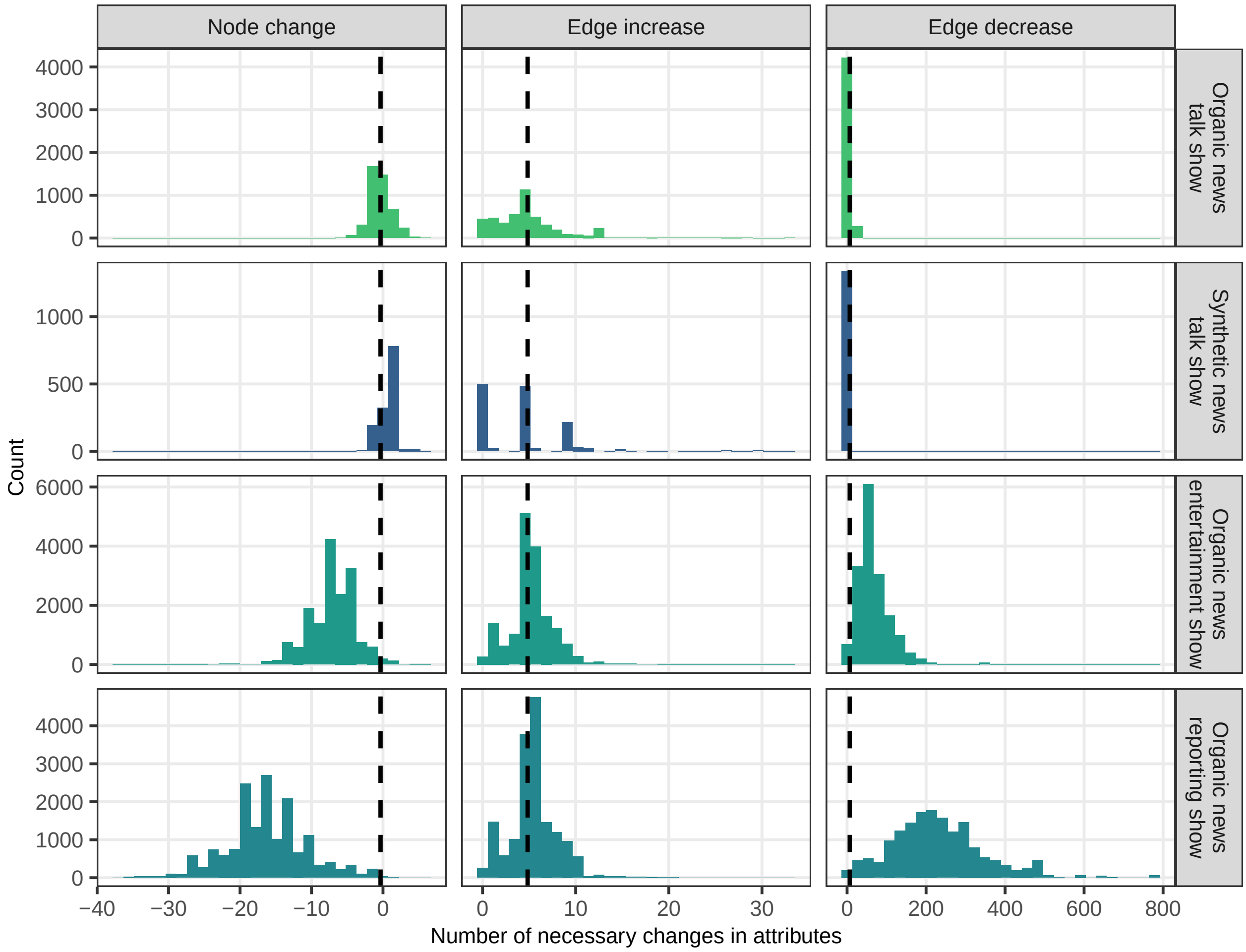

Figure 3: Comparison of ITSS systems in organic and synthetic news talk show transcripts, with benchmarks. The dashed lines mark the mean number of changes required when aligning the two samples of organic news talk shows (top row).

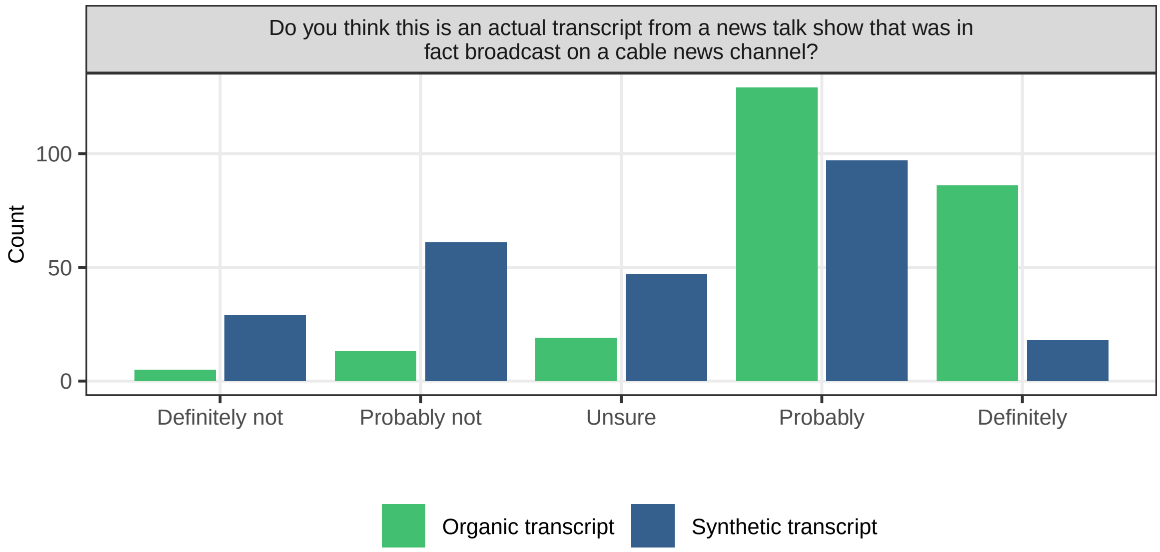

Figure 4: Responses to Question 1

do find it telling that only about 36% of participants thought the synthetic transcript was "probably" or "definitely" not an actual transcript from a broadcasted show.

Figure 5 displays the responses to Question 2. We see that the pattern of responses strongly supports H1: nearly three quarters (70%) of task participants said the synthetic transcript was either "somewhat similar" or "very similar" to what they expect a transcript of a news talk show broadcast on a conservative cable news channel to be like. (In comparison, 87% of participants thought the *Fox & Friends* transcript was "somewhat" or "very" similar.) Overall, the results of our assessment task – particularly the assessments in response to Question 2 – provide further evidence that the LLMs are adept at "naturally" generating cable news talk show transcript, or, more generally, documents rendering interior social worlds based on actors and their semantic meanings.[20] They also align with recent work showing that LLMs are capable of producing news reports that human readers deem as credible as human-written stories (Kreps et al., 2022).

5.2. How real synthetic transcripts affect perceptions of conservative actors and their arguments

Our second set of results shows how the synthetic transcripts can affect people's perception of the social world, specifically conservative actors, their ideas, and their programming. Figure 6 displays how participants' replies to Question 3 changed after reading a randomly assigned transcript type. We see that participants

[20]The assessment task's results also offer a validation of our semantic graph edit distance approach to comparison. Our conceptualization of the approach is that it captures valuable aspects of how actors and their relational meanings are portrayed in text. Since the results of the semantic graph edit distance indicate that the *Fox & Friends* and synthetic transcript are similar, then, if our approach is valid, people reading the two types of transcript should find them to be similar. This is what we see overall in the results of our assessment task (*i.e.*, Figure 4 and Figure 5).

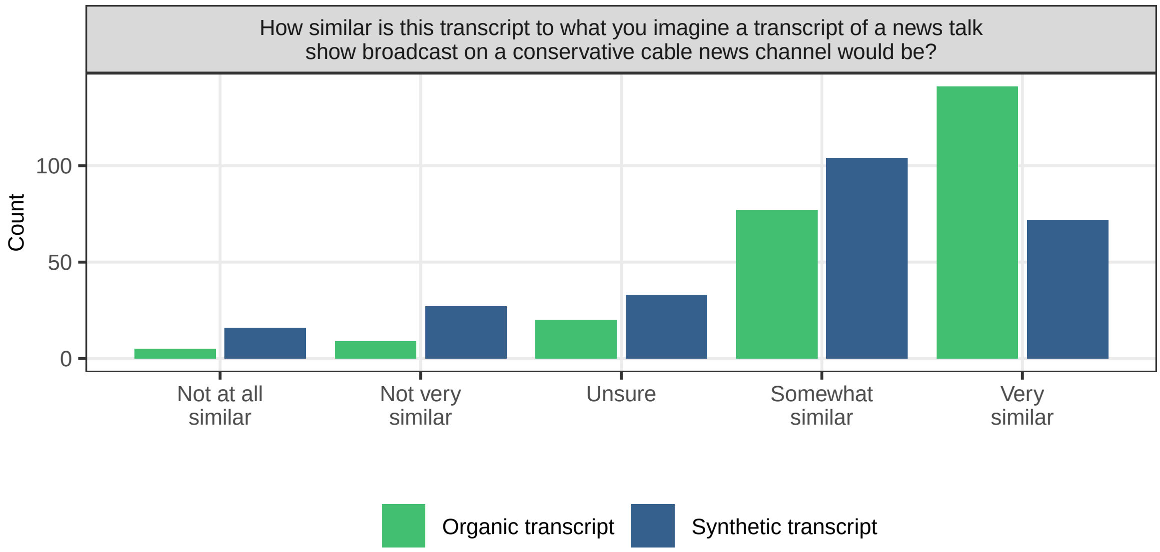

Figure 5: Responses to Question 2

who read the real synthetic transcript shifted their responses. On average, they came to see politically conservative viewpoints as more logical after reading the transcripts. This change is significant at the 0.05 level, although the effect size is modest, given that we recoded the responses an ordinal scale of 0 to 2. We also see that the manipulated synthetic transcripts nor the organic transcripts have an effect on participants' views on the logic of conservative arguments. Two aspects of these results – the modest effect size and organic transcripts' null effect – are important for understanding how the real synthetic transcripts may be influencing people's perceptions of information. We discuss these aspects in our Discussion section.

Figure 7 shows the change in responses to Question 4. The results are similar: on average, participants changed their replies to Question 4 after reading the real synthetic transcript, but not the organic or manipulated transcripts. That is, the readers of the real synthetic transcripts – and only these transcripts – understood Fox News programming as more convincing and influential (on the typical viewer they imagined) after reading the transcript. This change is significant at the 0.05 level and, once again, modest in size.[21] Once again, the effect size and null results are telling of potential underlying mechanisms. We discuss these in the following section.

6. Discussion

Our findings provide two insights about genAI and the texts it can produce. First, recent LLMs are adept at generating documents containing within their text social worlds based on actors and their relational

[21]We also recoded these responses on a scale from 0 to 2.

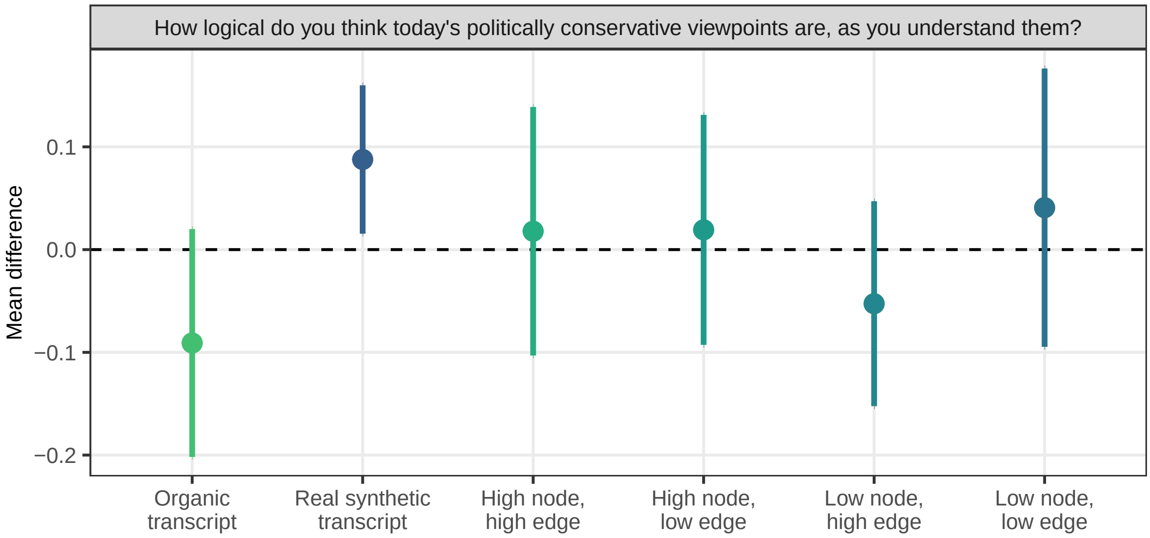

Figure 6: The effect of transcripts on responses to Question 3. Bars denote 95 percent confidence intervals.

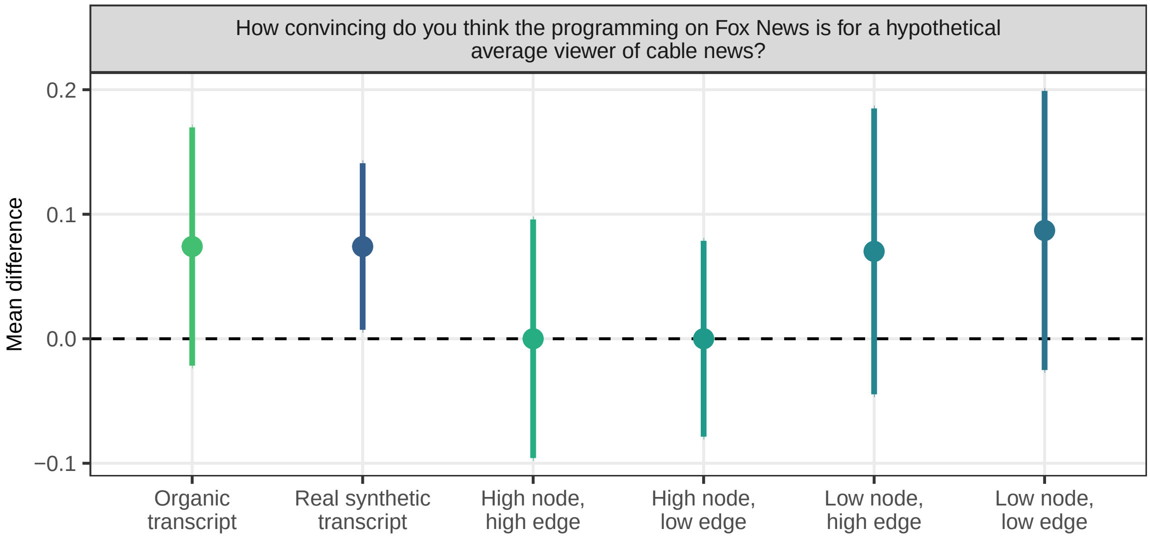

Figure 7: The effect of transcripts on responses to Question 4. Bars denote 95 percent confidence intervals.

semantic meanings (Lee and Martin, 2018). We consider the models adept because the output produced with minimal human guidance is similar to documents created by human-driven processes, such as the text transcriptions of television shows. This similarity, we argue, is in part due to the construction and rendering of social-semantic networks, or what we have referred to as ITSS systems. Second, the LLM documents produced "naturally" can influence people's perception of actors and their social meanings. Generated documents resulting from our manipulation of ChatGPT's inclinations did not affect people's perceptions.

The findings also present a puzzle. If our unmanipulated genAI texts can influence people's perceptions because they are good at embedding reasonable and realistic socio-semantic structures in text, then we should have found that the transcripts of true *Fox & Friends* broadcasts also affected people's perceptions. After all, these organic transcripts *do* encompass real socio-semantic structures. Yet, we did not find that the organic transcripts caused a change in how people understood conservative arguments and programming (see Figure 6 and Figure 7). Thus, our synthetic documents may differ from the organic documents in a way other than their "virtual social worlds".

If exchanges' social and topical, or content, dimensions are similar across organic and synthetic transcripts (compare, for example, Figure 1 and Figure 2), one way the document types might differ is in their style. That is, if the actors' relations to one another, the topics they talk about, and what semantic meanings they evoke in the texts are constant across transcript types, perhaps the types differ by *how* actors are portrayed as speaking. To explore this possibility, we conducted an exploratory analysis of the rhetorical style in the organic and real synthetic news talk show transcripts.[22] We found marked differences.

As shown in Figure 8, the real synthetic transcripts adopt an explicit-and-emotional style more often than the *Fox & Friends* transcripts. (This is the case for both the transcripts study participants read in the response task, seen in the left panel of Figure 8, and all the transcripts, seen in the right panel.) In the synthetic transcripts, speakers more frequently explicitly acknowledge their understanding (*e.g.*, "I understand your point"), explicitly state agreement (*e.g.*, "I agree with that") or disagreement (*e.g.*, "I do not agree with that"), and give "bare" commands, using unconjugated verbs at the start of sentences (*e.g.*, "Explain that for me"). They also use more negative and positive emotion words as descriptors (*e.g.*, "That was a good/bad point.").[23] We interpret the stylistic combination of explicitness and emotional terms as a kind of relatable, generic, and homogenized way of talking: it is unsubtle – explicit – in stating the point while also using familiar concepts – emotions – to describe things. Indeed, we also see in Figure 8 that the organic transcripts exhibit higher numbers of slightly more complex speech patterns, including giving agency,

[22]To conduct our analysis, we used the `politeness R` package (0.9.2).
[23]The stylistic categorizations and examples are adapted from Yeomans et al. (2018).

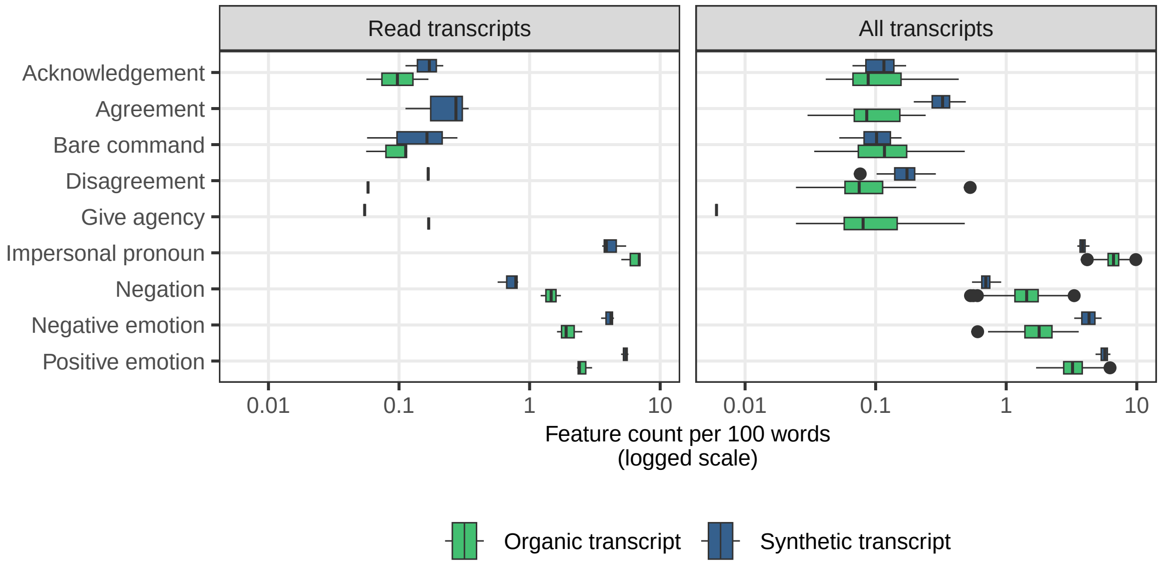

Figure 8: Stylistic features across transcript types

or suggesting actions for others (*e.g.*, "I want you to have an opportunity to clarify"), using impersonal pronouns (*e.g.*, "That is a good idea"), and negation, or using contradiction words (*e.g.*, "That cannot be your main point.").

The tendency for genAI transcripts to render actors' speech in a more relatable and homogenized style, relative to the true transcripts, raises two questions. First, why would this stylistic difference matter for how people perceive the social world? We posit that the more familiar and generic style of genAI transcripts make distasteful content more palatable. In other words, the real synthetic transcript shifted readers' views of conservative content by giving them versions of actors who were milder and more accessible, in terms of speaking style, than the actual people on *Fox & Friends*. This proposition aligns with Argyle, et al.'s (2023a) finding that participants in online conversations mediated by a ChatGPT-3 chatbot reported that those conversations had improved quality and tone.[24]

If our proposition is correct, then we would likely observe a change in views among the study participants who are not strong partisans – the people who can still update their impression of conservative rhetoric's logic and persuasiveness. This is why the modest effect sizes mentioned in Section 5.2 are telling. While the effects of the real synthetic transcripts are statistically significant (Figure 6 and Figure 7), their magnitudes are small because most of the partisans who read these transcripts did not change their views on conservative information. Instead, the people who changed were a handful of more moderate study participants who

[24] Our proposition also aligns with Spitale et al.'s (2023, 4) idea that GPT-3 may convey information more efficiently because it generates "text that is easier to read and understand compared to text written by humans."

shifted their answers slightly; most of the participants who responded to the synthetic transcript treatment self-identified as “moderate/middle of the road”, “slightly conservative”, or “liberal”.

The second question is, why might LLMs such as ChatGPT 3.5-Turbo depict social and discursive interactions in a more homogenized way? We believe the answer primarily lies in two aspects of how these LLMs are built. First, ChatGPT 3.5-Turbo and similar LLMs are parametric, meaning that they store information learned from a massive training dataset within the model and retrieve this information when generating output. Second, they have been fined-tuned with reinforcement learning from human feedback (RLHF), meaning that they have been trained beyond preceding ones (used as base models) to produce output that is better aligned with with the preferences of the people who provided the reinforcement (Ouyang et al., 2022; Ye et al., 2023). As a result of these two aspects, when we asked the model to generate actors and their speech, it drew from a huge expanse of information about talk show hosts, guests, and speakers, then sought to generate actor interactions and speech acts that it had learned would likely be acceptable to the people who participated in the RLHF process (*i.e.*, OpenAI researchers and contractors) (Ouyang et al., 2022).[25] , [26] The output thus skewed toward a “meaned”, or generic, depiction of a conservative talk news show that was more basic and plain than it otherwise would have been.[27]

7. Implications and conclusion

The answers to our questions about why stylistic differences would matter and why some LLMs would depict social and discursive interactions in simpler and more homogeneous ways point towards a few broader insights. First, for genAI and the concern among many model developers, policy makers, and others about LLMs’ ability to produce untruthful and toxic content: while a valid concern, our findings suggest that we should also be sensitive to *how* content is delivered. LLM-generated content that is explicitly toxic, misleading, or controversial may be upsetting to many, but it may ultimately not change people’s existing beliefs because it is easy to recognize. In contrast, as we have discussed, if some LLMs have a tendency to

[25]Of course, as Bender et al. (2021) point out, the voluminous information the model draws on reflects hegemonic groups, ideas, and structures in society; marginalized ideas and populations are unlikely to be represented in these data (see also Gillespie, 2024). This insight, however, is analogous to our point: the data that models draw on, as well as their design, could reinforce an “averaged” depiction of society, or “homogeneous hegemony”.

[26]To better understand the impact of synthetic style due to, in part, RLHF, we build on the comment of a reviewer and recommend that future research compare the influence of texts generated with models trained using different RLHF and/or with the signature of RLHF significantly minimized, possibly through prompt engineering or the fine-tuning of an existing model. This way, we could differentiate the effect of the default synthetic style of popular LLMs from the effects of LLMs, in general, as well as the style (or lack thereof) of models closer to the underlying base models of popular LLMs (*i.e.*, what exists before the RLHF training stage).

[27]Some LLM developers consider techniques like RLHF as opportunities to better align models with widely held normative values (*e.g.*, Anthropic, 2023), and we generally agree with this view. However, it is useful to consider the unanticipated – not simply unintended – consequences (de Zwart, 2015), such as more bovine content entering cultural circulation.

express politicized and ideological content in a more familiar, generic, and accessible style, it could be this stylistic tendency that gives synthetic content the potential to shape our shared social reality. This point leads to our second broader implication: ideas, opinions, arguments, and even certain political and media personalities – rendered in text by LLMs – that many people now dismiss out of hand could increasingly become more acceptable, and thus more broadly influential.

Another implication is methodological. Our study has illustrated how genAI documents can be used as tools in conducting social science research, rather than objects of research. Future work could continue exploring the potential for genAI – not only texts, but also images and videos – to be incorporated into research designs (see Jensen et al., 2022; Bail, 2023; Davidson, 2023). For example, synthetic cultural forms could be compared to organic ones to help uncover distinct features of the human and social creation and conceptualization of culture. Alternatively, organic cultural forms could be imputed into LLMs and systematically manipulated to create alternative (synthetic) versions that could plausibly have been created by people, and these alternative versions could then be used as part of a computational abductive exercise to advance theory (Karell and Freedman, 2019).

Finally, we conclude with remarks on the concept of duality. Duality – in both its theoretical articulation by Simmel (1955) and Breiger's (1974) analytical formalization – highlights how the intersection of groups within a person plays a role in determining individuality. That is, an individual is in part an expression of the combination of the social groups to which they belong, which are a subset of all available social groups. Our findings suggest that LLMs' depiction of people may increasingly unsettle this conceptualization of individuals and individuality. As parametric models trained on enormous collections of data, they have the ability – and the default tendency – to construct representations of people that are informed by a vast number of social groups. Put in terms of Simmel and Breiger, LLMs will produce texts expressing individuals as the intersection of nearly all social groups, or at least a previously unrealistic subset of possible social groups. The results are more generic versions of people.[28] Consequently, like how Simmel's ideas about intersecting circles and individuality grew out of his observations of modernity, today's social scientists may have to begin developing new conceptual tools to describe a growing homogeneity in our emerging world.

[28] Or, as mentioned in Footnote 25, because the groups in the unrealistically large subset will disproportionately be drawn from the hegemonic ideas, populations, and structures of society (Bender et al., 2021; Gillespie, 2024), people will be depicting as "generically hegemonic".

Acknowledgements

We thank our two anonymous reviewers and Andrew Davis, Omar Lizardo, and Kyle Puetz, the guest editors of the special issue. We also thank Ulrik Brandes, Ronald Breiger, John Levi Martin, Garry Robins, Tom A.B. Snijders and the other participants of the Duality@50 workshop, organized by Ulrik Brandes, Alessandro Lomi, Christoph Stadtfeld, and Natalie Peyer, for their comments and suggestions. We are indebted to Thomas Davidson, who provided helpful feedback during early stages of this project.

References

Alberto Acerbi and Joseph M. Stubbersfield. Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences*, 120(44):e2313790120, 2023.

Anthropic. Collective constitutional ai: Aligning a language model with public input. *Anthropic Research*, October 17: https://www.anthropic.com/index/collective–constitutional–ai–aligning–a–language–model–with–public–input, 2023.

Lisa P. Argyle, Christopher A. Bail, Ethan C. Busby, Joshua R. Gubler, Thomas Howe, Christopher Rytting, Taylor Sorensen, and David Wingate. Leveraging ai for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, 120(41):e2311627120, 2023a.

Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, page 10.1017/pan.2023.2, 2023b.

Sanjeev Arora, Yuanzhi Li, Yingyu Lian, Tengyu Ma, and Andrej Risteski. Linear algebraic structure of word senses, with applications to polysemy. *Transactions of the Association for Computational Linguistics*, 6:483–495, 2018.

Alina Arseniev-Koehler. Theoretical foundations and limits of word embeddings: What types of meaning can they capture? *Sociological Methods & Research*, DOI:10.1177/00491241221140142, 2022a.

Alina Arseniev-Koehler. Theoretical foundations and limits of word embeddings: What types of meaning can they capture? *Sociological Methods & Research*, page 10.1177/00491241221140142, 2022b.

Christopher Bail. Can generative artificial intelligence improve social science? *SocArXiv*, page https://osf.io/preprints/socarxiv/rwtzs, 2023.

Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv*, page https://arxiv.org/abs/2302.04023, 2023.

Nikita Basov. The ambivalence of cultural homophily: Field positions, semantic similarities, and social network ties in creative collectives. *Poetics*, 78(February):101353, 2020.

Nikita Basov and Julia Brennecke. Structure, content, and meaning in organizational networks: Extending network thinking. chapter Duality Beyond Dyads: Multiplex Patterning of Social Ties and Cultural Meanings, pages 87–112. Emerald Publishing, 2017.

Nikita Basov, Ronald Breiger, and Iina Hellsten. Socio-semantic and other dualities. *Poetics*, 78(February):101433, 2020.

Emily M. Bender, Angelina McMillan-Major, Timnit Gebru, and Shmargaret Scmitchell. On the dangers of stochastic parrots: Can language models be too big? *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, 2021.

Edward Bernays. *Propaganda.* 1928.

James Bisbee, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson. Synthetic replacements for human survey data? the perils of large language models. *Political Analysis*, page 10.1017/pan.2024.5, 2024.

Bart Bonikowski, Yuchen Luo, and Oscar Stuhler. Politics as usual? measuring populism, nationalism, and authoritarianism in u.s. presidential campaigns (1952-2020) with neural language models. *Sociological Methods & Research*, 51(4), 2022.

Ronald L. Breiger. The duality of persons and groups. *Social Forces*, 53(2):181–190, 1974.

Kathleen Carley. Extracting culture through textual analysis. *Poetics*, 22(4):291–312, 1994.

Youngjin Chae and Thomas Davidson. Large language models for text classification: From zero-shot learning to fine-tuning. *Open Science Foundation*, page https://osf.io/sthwk/download/?format=pdf, 2023.

Thomas Davidson. Start generating: Harnessing generative artificial intelligence for sociological research. *SocArXiv*, page https://osf.io/preprints/socarxiv/u9nft, 2023.

Frank de Zwart. Unintended but not unanticipated consequences. *Theory & Society*, 44:283–297, 2015.

Jacques Ellul. *Propaganda: The Formation of Men's Attitudes*. Vintage, 1973.

Stefan Feuerriegel, Renée DiResta, Josh A. Goldstein, Srijan Kumar, Philipp Lorenz-Spreen, Michael Tomz, and Nicholas Pr/"ollochs. Research can help to tackle ai-generated disinformation. *Nature Human Behavior*, 7:1818–1821, 2023.

Harry G. Frankfurt. *On Bullshit*. 2005.

Jan Fushe. *Social Networks of Meaning and Communication*. 2022.

Jan Fushe, Oscar Stuhler, Jan Riebling, and John Levi Martin. Relating social and symbolic relations in quantitative text analysis. a study of parliamentary discourse in the weimar republic. *Poetics*, 78(February):101363, 2020.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. Chatgpt outperforms crowd-workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120, 2023.

Tarleton Gillespie. Generative ai and the politics of visibility. *Big Data & Society*, 11(2):10.1177/20539517241252131, 2024.

Josh A. Goldstein, Jason Chao, Shelby Grossman, Alex Stamos, and Michael Tomz. How persuasive is ai-generated propaganda? *PNAS Nexus*, 3(2):1–7, 2024.

Kobi Hackenburg and Helen Margetts. Evaluating the persuasive influence of political microtargeting with large language models. *Proceedings of the National Academy of Sciences*, 121(24):e2403116121, 2024.

Geoffrey Hinton, Yoshua Bengio, Demis Hassabis, Sam Altman, and many others. Mitigating the risk of extinction from ai should be a global priority alongside other societal-scale risks such as pandemics and nuclear war. *Center for AI Safety*, pages https://www.safe.ai/statement–on–ai–risk#open–letter, 2023.

Maurice Jakesch, Advait Bhat, Daniel Bushek, Lior Zalmanson, and Mor Naaman. Co-writing with opinionated language models affects users' views. *arXiv*, page https://arxiv.org/abs/2302.00560, 2023.

Jeffrey Jensen, Daniel Karell, Cole Tanigawa-Lau, Nizar Habash, Mai Oudah, and Dhia Fairus Shofia Fani. Language models in sociological research: An application to classifying large administrative data and measuring religiosity. *Sociological Methodology*, 52(1), 2022.

Danel Karell and Jeffrey Sachs. How symbols influence social media discourse: An embedding regression analysis of trump's return to twitter. *SocArXiv*, https://osf.io/preprints/socarxiv/vz5rc, 2023.

Daniel Karell and Michael Freedman. Rhetorics of radicalism. *American Sociological Review*, 84(4):8726–753, 2019.

Daniel Karell and Michael Freedman. Sociocultural mechanisms of conflict: Combining topic and stochastic actor-oriented models in an analysis of afghanistan, 1979–2001. *Poetics*, 78(February):101403, 2020.

Daniel Karell, Michael Freedman, and Noam Gidron. Analyzing text and images in digital communication: The case of securitization in american white supremacist online discourse. *Socius*, 9, 2023.

Mickhail Khodak, Nikunj Saunshi, Yingyu Liang, Tengyu Ma, Brandon M. Stewart, and Sanjeev Arora. Proceedings of the 56th annual meeting of the association for computational linguistics. chapter A La Carte Embedding: Cheap but Effective Induction of Semantic Feature Vectors, pages 12–22. Association for Computational Linguistics, 2018.

Sarah Kreps, R. Miles McCain, and Miles Brundage. All the news that's fit to fabricate: Ai-generated text as a tool of media misinformation. *Journal of Experimental Political Science*, 9(1):104–117, 2022.

Kevin Lanning, Geoffrey Wetherell, Evan A. Warfel, and Ryan L. Boyd. Changing channels? a comparison of fox and msnbc in 2012, 2016, and 2020. *Analysis of Social Issues and Public Policy*, 21(1):149–174, 2021.

Mortiz Laurer, Wouter van Atteveldt, Andreu Casas, and Kasper Welbers. Less annotating, more classifying: Addressing the data scarcity issue of supervised machine learning with deep transfer learning and bert-nli. *Political Analysis*, page 10.1017/pan.2023.20, 2023.

Gaël Le Mens, Balázs Kovás, Michael T. Hannan, and Guillem Pros. Uncovering the semantics of concepts using gpt-4. *Proceedings of the National Academy of Sciences*, 120(49):e2309350120, 2023.

Monica Lee and John Levi Martin. Doorway to the dharma of duality. *Poetics*, 68(June):18–30, 2018.

Ryan Light and Colin Odden. Managing the boundaries of taste: Culture, valuation, and computational social science. *Social Forces*, 96(2):877–908, 2017.

Benjamin Lira, Margo Gardner, Abigail Quirk, Cathlyn Stone, Arjun Rao, Lyle Ungar, Stephen Hutt, Louis Hickman, Sidney K. D'Mello, and Angela L. Duckworth. Using artificial intelligence to assess personal qualities in college admissions. *Science Advances*, 9:eadg9405, 2023.

Kevin R. McKee, Andrea Tacchetti, Michiel A. Bakker, Jan Balaguer, and Lucy Campbell-Gillingham. Scaffolding cooperation in human groups with deep reinforcement learning. *Nature Human Behavior*, 7:1787–1769, 2023.

Paul McLean. *Culture in Networks*. Wiley, 2017.

Cade Metz. How could a.i. destroy humanity? *The New York Times*, June 10: https://www.nytimes.com/2023/06/10/technology/ai–humanity.html, 2023.

Raphaël Millière. Welcome to the next level of bullshit. *Nautilus*, October 11:https://nautil.us/welcome–to–the–next–level–of–bullshit–237959/, 2023.

Ann Mische. The sage handbook of social network analysis. chapter Relational Sociology, Culture, and Agency, pages 80–97. Sage, 2011.

Brent Mittelstadt, Sandra Wachter, and Chris Russell. To protect science, we must use llms as zero-shot translators. *Nature Human Behavior*, 7:1830–1832, 2023.

John Mohr. Measuring meaning structures. *Annual Review of Sociology*, 24:345–370, 1998.

Fabio Motoki, Valdemar Pinho Neto, and Victor Rodrigues. More human than human: Measuring chatgpt political bias. *Public Choice*, 198:3–23, 2023.

Laura K. Nelson. Leveraging the alignment between machine learning and intersectionality: Using word embeddings to measure intersectional experiences of the nineteenth century us south. *Poetics*, 88:101539, 2021.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almedia, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christian, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. *arXiv*, page https://arxiv.org/abs/2203.02155, 2022.

Mark Pachucki and Ronald Breiger. Cultural holes: Beyond relationality in social networks and culture. *Annual Review of Sociology*, 36:205–224, 2010.

Emma Rodman. A timely intervention: Tracking the changing meanings of political concepts with word vectors. *Political Analysis*, 28(1):87–111, 2020.

Pedro Rodriguez, Arthur Spirling, and Brandon M. Stewart. Embedding regression: Models for context-specific description and inference. *American Political Science Review*, DOI:10.1017/S0003055422001228, 2023.

Jasper Roe and Mike Perkins. 'what they're not telling you about chatgpt': Exploring the discourse of ai in uk news media headlines. *Humanities and Social Sciences Communications*, 10:10.1057/s41599–023–02282–w, 2023.

Kevin Roose. Inside the white-hot center of a.i. doomerism. *The New York Times*, July 11: https://www.nytimes.com/2023/07/11/technology/anthropic–ai–claude–chatbot.html, 2023.

Camille Roth and Jean-Philippe Cointet. Social and semantic coevolution in knowledge networks. *Social Networks*, 32:16–29, 2010.

Alix Rule, Jean-Philippe Cointet, and Peter S. Bearman. Lexical shifts, substantive changes, and continuity in state of the union discourse, 1790–2014. *Proceedings of the National Academy of Sciences*, 112(35):10837–10844, 2015.

Bruce Schneier and Nathan Sanders. Who's accountable for ai usage in digital campaign ads? right now, no one. *Ash Center for Democratic Governance and Innovation*, October 11:https://ash.harvard.edu/who2023.

Georg Simmel. *Conflict and The Web of Group-Affiliations*. The Free Press, 1955.

Giovanni Spitale, Nikola Biller-Andorno, and Federico Germani. Ai model gpt-3 (dis)informs us better than humans. *Science Advances*, 9:eadh1850, 2023.

Jason Stanley. *How Propaganda Works*. 2016.

Dustin S. Stoltz and Marshall A. Taylor. Cultural cartography with word embeddings. *Poetics*, 88:101567, 2021.

Oscar Stuhler. Who does what to whom? making text parsers work for sociological inquiry. *Sociological Methods & Research*, 51(4):1580–1633, 2022.

Taylor Webb, Keith J. Holyoak, and Hongjing Lu. Emergent analogical reasoning in large language models. *Nature Human Behavior*, 7:1527–1541, 2023.

Sam Woolley and Phil Howard. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. 2017.

Junjie Ye, Xuanting Chen, Nuo Xu, Can Zu, Zekai Shao, Schichun Liu, Yuhan Cui, Zeyang Zhou, Chao Gong, Yang Shen, Jie Zhou, Simin Chen, Tao Gui, Qi Zhang, and Huang Xuanjing. A comprehensive capability analysis of gpt-3 and gpt-3.5 series models. *arXiv*, page https://arxiv.org/pdf/2303.10420.pdf, 2023.

Michael Yeomans, Alejandro Kantor, and Dustin Tingley. The politeness package: Detecting politeness in natural language. *The R Journal*, 10(2):489–502, 2018.

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. Can large language models transform computational social science? *arXiv*, page https://arxiv.org/abs/2305.03514, 2023.