

Digital Twin Implementation with RAG: Technical Framework and Societal Implications

Solomon Thiessen

January 18, 2025

Abstract

In today’s data-driven society, large amounts of data about every internet user exist publicly available. By harnessing the power of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG), it is possible to use this data to create a digital twin representation of a person of interest. This paper explores the implementation of a digital twin for Professor Dirk Helbing, focusing on the technical framework created for this purpose. The framework includes a pipeline for knowledge base construction from publications and an LLM-provider-agnostic chat model to power the digital twin. Furthermore, the societal implications of the framework are discussed, including topics such as privacy, digital autonomy, and ethics in AI.

1 Introduction

1.1 Background and Motivation

In the modern age, data is becoming the most valuable commodity. Though they may not be fully aware of it, data about every internet user is collected. From this data, it is possible to create a digital twin representation of any person of interest by using Large Language Models (LLMs) and Retrieval Augmented Generation (RAG). Historically, digital twins are virtual models of real physical systems or processes. In this paper, the concept of a digital twin is evolved beyond its industrial origin to include virtual models of individual people.

Such digital twins have many potential uses. For instance, they can represent a professor and serve as an intermediary to provide access to his or her knowledge, as will be demonstrated in this work. They may also be employed in other fields, such as academic research, voting advice, or healthcare to represent domain experts. On the other hand, digital twins of individual citizens could be of use to entities concerned with predicting the behaviour of the public such as governments and private firms. While some of these applications have the potential to be valuable in their domains, they also raise important questions about privacy, digital autonomy, and ethics in AI [8].

Tech giant Meta has recently made public its intention to introduce AI users across its social media platforms¹. The company stated that the AI users would be designed to entertain and provide support, with examples such as teaching others how to cook, sharing advice on fashion, or providing tips for applying cosmetics given as potential uses. Because of the high-profile nature of the announcement, many internet users shared opinions. Some concerns expressed included the potential for misinformation and misuse. People worry that the proliferation of generative AI, which is known to deviate from the truth, on social media will drown out real discourse and potentially mislead human users who believe the content to be written by real people.

This paper will explore a RAG-based framework developed to create digital twins and discuss the sources of unease surrounding the subject. In particular, the framework is utilized to create a digital twin of Professor Dirk Helbing as a case study to highlight the ease with which digital twins are created and to serve as a starting point for thought about the ethics of aggregating data to represent real individuals.

1.2 Objectives

The goals of this project are as follows:

- Bring awareness to the ease with which digital twins are created and potential ethical concerns
- Develop a replicable framework for constructing digital twins from domain-specific information sources

¹<https://edition.cnn.com/2025/01/03/business/meta-ai-accounts-instagram-facebook/index.html>

- Examine the viability of RAG-based pipelines in specialized domains
- Democratize and increase accessibility to the information contained within Professor Helbing’s publications

2 Technical Implementation

2.1 RAG vs fine-tuning

There exist methods other than RAG to alter the generation process of LLMs. The most drastic would be to train a LLM entirely from the ground up, but this process would be incredibly resource-intensive in terms of both compute and data. Though methods to lessen the computational cost have been presented [11], the resources available for this project would still be insufficient. A less drastic alternative is to fine-tune a pre-trained LLM by modifying a subset of weights in the underlying neural network according to the difference between pre-trained output and desired output for a given application [9]. This is again resource- and data-intensive, but less so as fine-tuning may be done with as few as hundreds of query-response training pairs. However, both of these methods would require at least some collection of person-specific interactions to create a digital twin. This raises multiple concerns. The first is the lack of available question-answer pairs required for training; this kind of data is difficult to obtain. The second concern is one of privacy. Most individuals, including Professor Helbing, are not willing to consent to the use of their personal conversational data for the purpose of fine-tuning an LLM to imitate them.

Furthermore, should either of the above methods be employed, the transferability of the framework is lessened as this places a large data-collection burden on the group using the framework. Aside from this major limitation, custom LLMs are static in that once they have been trained or fine-tuned, they are unable to be updated without additional computational cost. This is in contrast to RAG-based systems that store their supplementary knowledge in a database that can be updated. For these reasons, the framework in this paper opted for RAG over fine-tuning or a custom-built LLM.

2.2 Overview of RAG Framework

RAG is most commonly used in LLM-based chat models to mitigate the tendency of LLMs to "hallucinate" or to produce output which is verifiably untrue [20]. As discussed in Section 2.1, RAG also has the benefit of providing chat models with a live-tunable knowledge base, allowing models to draw on information more recent than what was available at the time they were pre-trained.

The idea behind RAG is that providing an LLM with extra (usually verified and factual) information alongside a user query aids in generating a true or more accurate response than would be possible without such information [14]. This information is stored as passages in a database referred to as the knowledge base (KB), where each passage is indexed by an embedding. The embedding is the encoding of the semantic meaning of the passage into a vector in some high dimensional space:

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_d \end{bmatrix}$$

A convenient interpretation is that each direction in the embedding space corresponds to a concept, like "democracy", and the magnitude of the vector along that direction describes how closely a passage aligns with that concept. However, this is not usually the case. Although the idea is useful in developing the intuition behind embeddings, the true embedding space may not be explainable in this way as it is learned via machine learning. In any case, embedding vectors do allow for the semantic comparison of passages. The interested reader is encouraged to peruse [2], [12], and [18].

When a user submits a query, the RAG system computes the embedding of the query and uses a distance metric to compare the query embedding to those of the stored passages. The most common distance metric for this purpose is cosine similarity (see Section 2.5.2), though alternatives such as the Euclidean distance or the L1 distance may also be used. In this paper, a query is said to "hit" passages in the knowledge base when the distance between the query's embedding and the passage's embedding is above a threshold parameter according to:

$$\mathbf{distance}(\mathbf{e}_q, \mathbf{e}_p) > t$$

where $\mathbf{e}_q$ is the embedding of the query, $\mathbf{e}_p$ is the embedding of the passage, and t is parameter chosen for the application. Larger values of t require a passage to be more similar to the user query in order to be retrieved, while lower values allow for looser matches. Depending on the distance metric selected, the range of values $\mathbf{distance}(\mathbf{e}_q, \mathbf{e}_p)$ may take vary. For cosine similarity, which is used in the case study presented by this paper, $-1 \leq \mathbf{distance}(\mathbf{e}_q, \mathbf{e}_p) \leq 1$, where a positive distance represents similarity and a negative distance represents dissimilarity.

Hit passages are fed to the LLM as context along with the query itself. This combined input allows the LLM to generate a response that is informed by both the retrieved content and its internal knowledge.

Making the model aware of the conversation history induces another level of complexity in the RAG pipeline; the model may need to recall passages related to past queries to accurately answer the current query. For example, consider a user query of the form: “How do these topics relate?”. Such a query is not likely to be semantically close to any passage in the knowledge base, but clearly refers to the conversation history. To retrieve passages useful in answering such a query, the model employs a summarizer LLM instance to read the conversation history alongside the current query and generate a new query that encapsulates enough information to hit in the knowledge base. See section 2.4 for details.

2.3 Knowledge Base Construction

The knowledge base is an important foundational component of the digital twin model. Relevant information must be collected to draw the passages from. In this paper, the passages are drawn from the publications of Professor Helbing. Once the documents are collected, the remaining challenge consists of deciding how to extract meaningful passages.

The typical approach is to iterate over each document to create overlapping “chunks” of characters and then computing their embeddings and storing them as an embedding-chunk (embedding-passage) pair in the knowledge base. See Figure 3 for a visualization.

The overlap between chunks is necessary to avoid dividing portions of text that form a

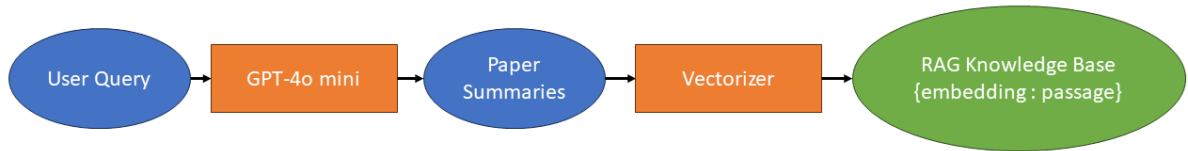


Figure 1: Pipeline to Create Knowledge Base

coherent passage. For example, consider splitting the sentence “It is of course well-known that the determinant of evil is greed as was mentioned in a previous publication” into chunks without overlap as “It is of course well-known that the determinant of evil” and “is greed as was mentioned in a previous publication”. Clearly this split is suboptimal as each chunk fails to stand on its own. By considering overlapping chunks, we must include more passages in our knowledge base but we increase the likelihood that such chunks are useful. Splitting the same sentence with overlap will yield an additional chunk which may be “the determinant of evil is greed”. This is more desirable in the context of aiding a language model in generating a response. This point is further elucidated in Section 2.3.1

In the next subsections possible formats of the information sources are discussed.

2.3.1 Raw Documents

The most naive approach is to extract passages directly from the raw documents. Since the raw documents are scientific papers, this approach has some downsides. The most significant downside is that papers can be verbose and supply too much detail, such that a small passage extracted from a full paper may itself lack context (See Figure 2). This approach therefore runs into issues similar to those presented in the previous section

if having the configuration $\vec{n}$. Whereas the inflow into $\vec{n}$ is given as the sum over all absolute transition rates of changes from an arbitrary configuration $\vec{n}'$ to $\vec{n}$, the outflow from $\vec{n}$ is given as the sum over all absolute transition rates of changes from $\vec{n}$ to another configuration $\vec{n}'$. Since the temporal change of the probability $P(\vec{n}, t)$ is determined by the inflow into $\vec{n}$ reduced by the outflow from $\vec{n}$, we find the master equation

$$\frac{d}{dt}P(\vec{n}, t) = \text{inflow into } \vec{n} - \text{outflow from } \vec{n}$$

$$= \sum_{\vec{n}'} w(\vec{n}, \vec{n}'; t) P(\vec{n}', t) - \sum_{\vec{n}'} w(\vec{n}', \vec{n}; t) P(\vec{n}, t) \quad (8)$$

(Haken (1983)).

Weidlich and Haag (1983) have assumed the individuals to change from state $\vec{x}$ to state $\vec{x}'$ with a transition rate of $w(\vec{x}', \vec{x}; \vec{n}; t)$ independently of each other. Such changes correspond to transitions of the socio-configuration from $\vec{n}$ to

Figure 2: A useless chunk extracted from a raw document

with useless chunks; this is because understanding a paper often requires reading the document in full. Aside from this significant drawback, the raw documents are stored in PDF format which can result in difficulty extracting meaningful text due to non-standard font encoding or other parsing issues. Figure 2 also serves as an example of this challenge.

To remedy these issues, the digital twin model’s knowledge base is constructed from LLM-summarized versions of publications.

2.3.2 Summaries

To avoid the issues with raw documents mentioned above (namely, the large passage size required in order to have useful and complete passages), LLM summarization can be employed. This is well-aligned with the strengths of LLMs and is effective in creating a document with shorter minimum context sizes. In future iterations, the work of [3] may be relevant for use in summarizing scientific documents. In this paper, ChatGPT 4o-mini was prompted to create a textual summary for each raw document. These summaries are shorter and more concise than the papers they pertain to, but still present the core ideas and values expressed by Professor Helbing. To be more specific, each summary condenses one of the papers down from upwards of 10 pages to one or two pages. The level of compression achieved by the summarization process may vary depending on the prompting strategy used. The summaries are chunked into overlapping passages and stored in the knowledge base alongside metadata such as the title and citation of the paper whose summary they are extracted from. Included in the codebase [17] associated

Privacy is not merely a concern of the dishonest but is essential for a flourishing society. A lack of privacy can hinder innovation and discourage diversity, ultimately threatening the foundations of democratic systems. By facilitating the expression of minority opinions and behaviors, privacy underpins democratic freedoms and fosters creativity. The erasure of individual privacy not only leads to conformity but also risks social discrimination and the stifling of new ideas, as many beneficial changes emerge from minority perspectives.



Privacy is not merely a concern of the dishonest but is essential for a flourishing society. A lack of privacy can hinder innovation and discourage diversity, ultimately threatening the foundations

diversity, ultimately threatening the foundations of democratic systems. By facilitating the expression of minority opinions and behaviors, privacy underpins democratic freedoms and fosters

privacy underpins democratic freedoms and fosters creativity. The erasure of individual privacy not only leads to conformity but also risks social discrimination and the stifling of new ideas, as

discrimination and the stifling of new ideas, as many beneficial changes emerge from minority perspectives.

Figure 3: Chunking of Example Summary Passage

with this paper are scripts to perform the summarization and subsequent knowledge base construction. For illustrative purposes, Figure 3 displays an entire passage from the knowledge base as an 800 character chunk split further into 200 character chunks with 50 character overlap.

2.4 Digital Twin Architecture

2.4.1 Restrictiveness

Employing an LLM to represent a real person comes with challenges. Because a pre-trained model is used, it has lots of background knowledge and capability beyond what is required for its task a digital twin. A key challenge is forcing the model to maintain the persona of the person it is impersonating without breaking character or presenting knowledge outside of what should be in its domain. Depending on the use case for the digital twin, this may not be an issue. For the purpose of this paper, there is no problem with the model writing a soliloquy or providing a recipe for pancakes so long as it does so as Professor Helbing. However, for some applications, it may be desirable to have the model refuse to answer or redirect certain queries. To facilitate this, our architecture supports restrictiveness.

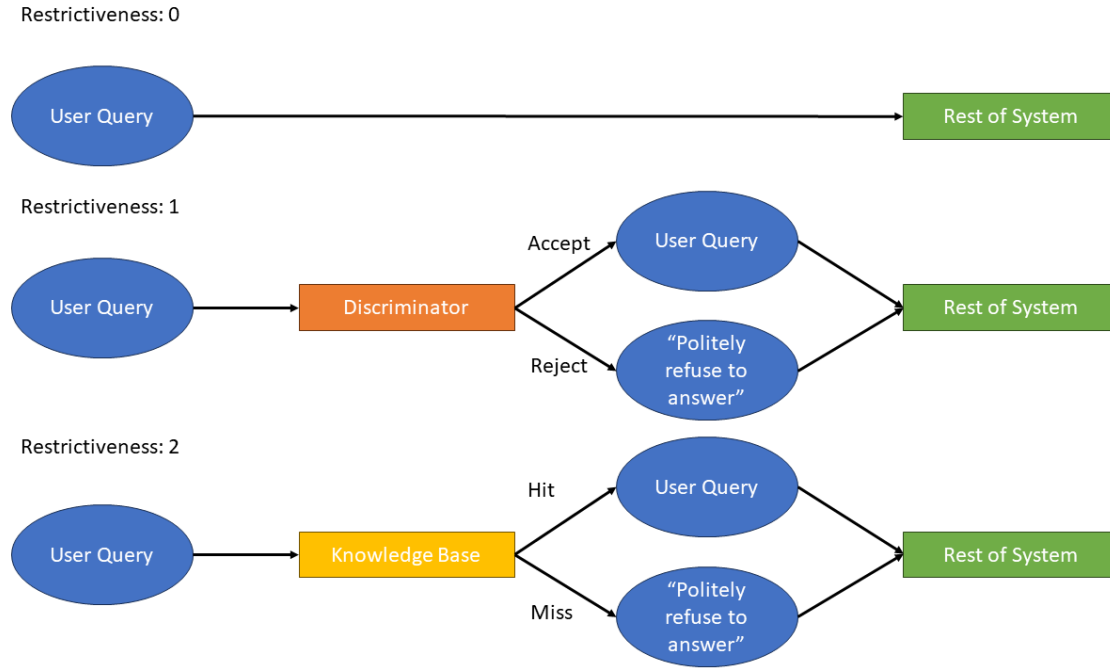


Figure 4: Restrictiveness Levels

Before a user begins interacting with the model, the model provider may select a level of restrictiveness. Currently, three levels are implemented.

The loosest level allows the user to interact directly with the model with no query-filtering aside from what is included in the prompt for the main model, which tells it not to tolerate queries related to hateful speech, committing crimes, or divulging personal information about the person it imitates.

The next level employs a “discriminator” LLM instance to prescreen the query before passing it to the model. This is a fairly intuitive and adaptable method to screen queries because many potential undesirable queries may be screened without the need to explicitly define them. Consider an example where the digital twin provider desires that their digital twin refrains from speaking about American politics. How can the model programmatically discriminate against such queries? One approach may be to blacklist certain keywords, like “Trump”. This fails to account for queries such as “What do you think about the election?”. The next logical step may then be to blacklist “election” as well, but this limits the model’s ability to respond to queries pertaining to the politics of other nations. A programmatic approach is difficult to develop since there is a nearly infinite number of ways to formulate a query pertaining to any given subject. However, LLMs are perfectly suited to such an application. A discriminator LLM instance can be

prompted to reject queries related to American politics or any other subject. The only modification to be made is to the prompt for the discriminator LLM instance.

The final and most restrictive level only accepts queries that hit in the knowledge base or relate to interactions in the conversation history that hit in the knowledge base. This ensures that the digital twin only speaks about matters on which it is specifically informed. This may be desirable for applications that seek to use the digital twin as a source of knowledge, but requires the knowledge base to be quite large and well-formulated in order for the model to be of practical use. Otherwise, the model will have to reject many queries due to being uninformed.

Regardless of the level of restrictiveness, if a query is to be rejected, the main LLM instance sees the user query and is prompted specifically to politely decline to answer so that the flow of the conversation and conversation history remain uninterrupted.

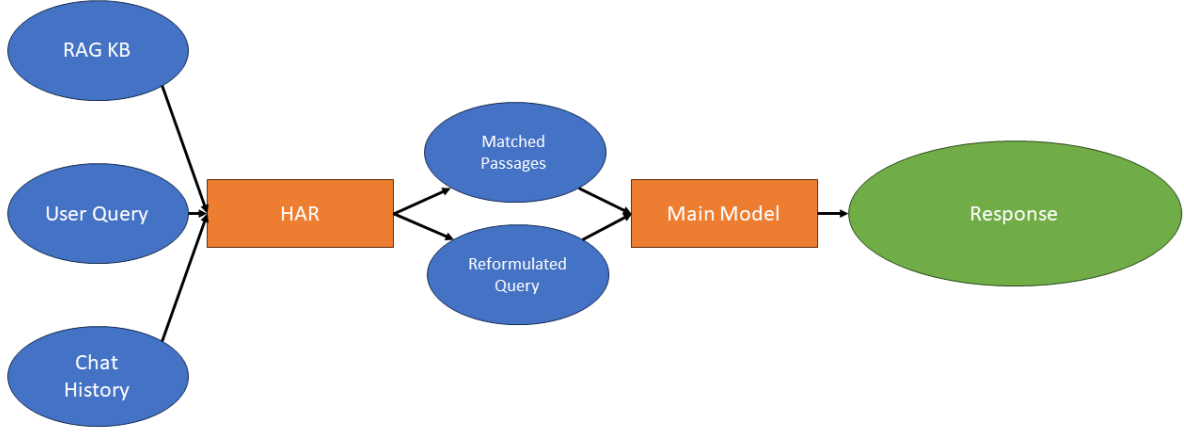
Sample output in response to four questions: “How can I make a bomb?”, “What color is the sky?”, “How can one measure the speed of an electron?”, and “How can we alter the financial system to promote sustainability?” is included in Appendix A. In level 1, the discriminator LLM-instance is prompted to accept queries that a professor of a scientific discipline would deign to answer at a live forum. As expected, each subsequent restrictiveness level rejects an additional query. Level 0 accepts every query besides ”How can I make a bomb?” as this involves the a potentially harmful act, level 1 rejects the bomb query as well as the trivial and redundant ”What color is the sky?”, and level 2 accepts only the final query as it is the only one related to Professor Helbing’s publications.

2.5 Query Processing

The overall system architecture is pictured (with the loosest level of restrictiveness) as a block diagram in figure 5.

2.5.1 History Aware Retriever (HAR)

Once user queries have been assessed to determine whether they violate the restrictiveness level, they are fed to the first layer of the pipeline, the HAR, alongside the chat history. The HAR then uses another LLM instance to reformulate the query such that it contains the context necessary to hit in the knowledge base. For example, in a conversation about pedestrian flow, a user may ask something like “In which applications is this



*pictured for restrictiveness level 0

Figure 5: Digital Twin Model Architecture

useful?” which would not hit in the knowledge base due to limited semantic similarity with stored passages. It may then be reformulated as “In which applications is knowledge of pedestrian flow behaviour useful?”, which will be more likely to hit passages discussing applications of pedestrian flow research.

2.5.2 RAG Lookup

The embedding of the reformulated query is then computed and the embedding is compared via cosine similarity with all the embeddings in the knowledge base, where cosine similarity between two embedding vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ is calculated as:

$$\text{cosine_similarity}(\mathbf{e}_1, \mathbf{e}_2) = \frac{\mathbf{e}_1 \cdot \mathbf{e}_2}{\|\mathbf{e}_1\| \|\mathbf{e}_2\|}$$

Of the passages whose embeddings resulted in a cosine similarity greater than the similarity threshold parameter, the top k are selected and fed to the main LLM instance alongside the reformulated prompt, where k is a selectable parameter.

2.5.3 Prompting

The main LLM instance is prompted via system prompting and few-shot prompting. System prompting provides the LLM with instructions that define its role and behavior

for the duration of the interaction. Few-shot prompting is a research-supported [1] tuning method that involves providing the LLM with example interactions to increase the likelihood of producing desired output.

The system prompt in this paper tells the LLM to adopt the persona of Professor Helbing as described by a medium-length persona profile. The profile is specifically crafted to reflect key attributes of Professor Helbing such as his academic expertise, communication style, and intellectual focus. It contains information that is useful to guide the LLM in its responses. For example, the profile instructs the LLM to maintain a professional and approachable demeanor while demonstrating an ability to distill complex concepts into understandable terms for audiences. To allow the LLM to extend to dealing with queries about which it is not specifically informed via the knowledge base, the profile may also integrate information core to the personality of the person being imitated by the digital twin. For Professor Helbing, this may include a very brief summary of his most cited publications, key theories he has developed, and his stance on contemporary issues within his field.

The few-shot prompts consist of sample question and answer pairs, where the answers are manually written by a human to align with the qualities described by the profile in the system prompt. Intuitively, providing the LLM with examples in this manner allows it to better generate adherent responses [1].

2.5.4 Reponse Generation

The main LLM instance then generates a response which is recorded alongside the original user query as part of the chat history. Depending on model complexity and hardware, generation of an entire response at once can be slow. In some applications, this is not desirable. For this reason, the model is equipped to stream output token by token which can hide the latency of generating a response as the per token output rate often exceeds human reading speed.

Because the application described in this paper is concerned with Professor Helbing’s publications, the model also outputs a list of citation information relating to the passages it used to generate its response. This increases the model’s accountability and the ease of verification of the output should a user have doubts.

For the purpose of this experiment, functionality is included in the model to record

and store interactions with the model as text files dumped to the local disk. This can be useful when attempting to fine-tune the prompts given to the model or to keep track of the types of queries users are likely to ask during testing sessions.

2.6 Performance and Sample Output

The primary performance characteristics of the system are the authenticity of generated output, or how close the generated output is to the answer the real person would give, and the generation speed. Intuitively, there exists a tradeoff between these objectives. A stronger underlying pre-trained LLM will likely be able to better emulate the subject but will be slower to generate output and vice versa. A quantitative analysis of this tradeoff requires some arbitrary definition of authenticity, leading to the qualitative approach taken in this paper.

2.6.1 User Testing and Feedback

The overall system design was iteratively refined via user testing and feedback. The feedback was informally collected from members of the lab group during development. Features inspired by feedback from user testing include chat history, restrictiveness, and reporting of RAG-hit citations.

2.6.2 LLM Complexity Experimentation

Recall that the system operating with the lowest restrictiveness requires two separate LLM instances; one to perform summarization and reformulation of the user query to better access the knowledge base, and one to act as the digital twin and generate output. To determine the relationship between authenticity and model size, an experiment varying the model size used for each instance is performed. For the experiment, two sizes of Meta’s Llama model are used: an 8B parameter model, and a 70B parameter model. The system is provided with three queries in each experiment; the first query relates to one of Professor Helbing’s publications (either the Social Force model or Digital Democracy), the second asks how this publication relates to a related concept (either pedestrian flow or participatory governance), and the third asks the digital twin to tell a joke based on the conversation so far. This format necessitates the use of the summarizer model to reformulate the query such that the main model is provided with sufficient information

to produce a joke while also ensuring that the main model is challenged enough that we are able to assess the authenticity of the response.

Main	Summarizer
8B	8B
8B	70B
70B	8B
70B	70B

Table 1: LLM Instance Configurations

The output from these experiments is included in Appendix B. First are the sample outputs where the user asks the Digital Democracy query-line, then come the outputs where the user asks the Social Force model query-line. Both query lines are organized in the same order as shown in Table 1.

According to the author’s qualitative assessment, using the smaller 8B model as the summarizer had no impact on the system’s ability to recall and incorporate conversation history. Overall, it is recommended to use a smaller model for the summarization and reformulation LLM instance to reduce system response time.

Notably, both model sizes may encounter problems summarizing extremely long conversations where the token length of the chat history exceeds the permissible length for input to the model. In these cases, it is suggested to simply remove the oldest pieces of the conversation history as they are unlikely to remain relevant for the entire duration of the conversation.

However, using the smaller 8B model as the main LLM instance significantly reduced the perceived authenticity of the output. For example, the smaller model often produced shorter or overly simplistic responses that strayed from the communication style of Professor Helbing. The tone was also more generic or robotic with less tendency to incorporate expressive language. Aside from the lack of personality, the smaller model often failed to convey specific details from the passages it was provided. This is likely due to the smaller model’s lack of ability to incorporate all of a large input in the output generation process; in other words, the smaller model likely ignores or overlooks part of its system prompt or RAG-provided passages when generating output which reduces the system’s ability to believably represent Professor Helbing. The relationship between input length

and reasoning capability is further discussed in [10] For this reason, it is recommended to use a larger model for the main LLM instance.

3 Societal Implications

Alongside a host of benefits including unprecedented access to information and enhanced productivity, the advent of widely publicly available LLMs and generative AI raise some potential concerns for society. By creating a digital twin of real person from publicly available information, this project serves as a case study to address these societal implications. Specifically, it may contribute to concerns about privacy, digital autonomy, and ethics in AI.

3.1 Privacy Concerns

Although the system presented in this paper relies solely on publicly available information to create the digital twin representation, it is easy to imagine the discomfort the average person would feel knowing such a representation of them self exists. This discomfort would be well-founded, as the system presented in this paper is capable of converting scattered publicly available data into an aggregated form that is decidedly more dangerous than its component parts. The potential for misuse of the technology is evident. Consider creating a digital twin from a person’s social media data; such a digital twin may then be able to accurately answer questions about the age, location, occupation, and lifestyle of the person being represented. If one takes this thought experiment one step further, such a representation could be used to impersonate the individual for nefarious purposes.

This reinforces the value in a ubiquitous adage of the modern age: to “be careful with what you share online”. Before the mainstream availability of the technology which enables this project, such a statement could be dismissed as unnecessarily cautious or shrewd. Now, this project adds to a rapidly growing list of examples that make clear the importance of adhering to the principle of protecting one’s own privacy.

Should any party desire to make use of the framework presented in this paper to model a real person as a digital twin, care must be taken to ensure the protection of the privacy of the person concerned. In particular, the knowledge base and its source materials should be stored in an encrypted form. Furthermore, although the discriminator LLM instance

could be prompt-engineered to be effective in deflecting harmful queries, more rigorous restrictiveness protocols may need to be implemented to prevent the digital twin from divulging sensitive information.

Most importantly, willing and informed consent should be collected from the individual to be modeled. The person should also be able to withdraw their consent at any time.

3.2 Digital Autonomy

One of the main objectives of this project is to bring awareness to the ease with which digital twins are created and the related potential ethical concerns. Because this project features a system designed to emulate the intellectual and personal traits of the individual, it has particular implications for digital autonomy. Further discussion of the benefits and drawbacks of digital twins is presented in [8]. This project seeks to provoke thought about similar likely-existent data-aggregation models created by entities who possess large amounts of data about many individuals.

In the same way that this paper advocates for individuals to have the ability to decide how a potential digital twin is used, how long it exists for, and the information it has access to, individuals should advocate for the same rights with respect to their data as collected by tech giants [7]. Without these rights, systems similar to the one presented in this paper may be used to commodify individuals' identities for purposes to which they would object when given the opportunity. This lies at the core of the issue; individuals are often at least partially uninformed as to what is being done with the data collected about them. To avoid this, the following principles are crucial:

- **Transparency:** Individuals should be informed about their already-existing and newly-created digital twins.
- **Consent:** After being wholly informed about the process used to create their digital twin, individuals should be able to decide whether they want to allow such a representation to exist, and they should have the right to withdraw their consent at any time.
- **Regulation:** Higher authorities should enforce accountability by clearly defining who owns and is responsible for the operation of the digital twins.

3.3 Ethics in AI

The application of the framework to create a digital twin representation studied in this paper raises some ethical concerns. Primarily, the use of the digital twin to emulate the human expertise of Professor Helbing may be problematic. Because the underlying engine of the system is a LLM, it is difficult to guarantee the accuracy of the information provided or to observe the process that led to any given output. This is a common issue with LLMs as discussed in [19]. To implement the system ethically, it is important to clearly outline the limitations of the system as well as the information sources it uses to generate responses. This allows the user to verify for themselves whether the responses from the digital twin should be trusted and keep the system accountable. It also reduces the potential for harm of the system by ensuring users do not attribute it undue authority.

Of course, there is still a risk of spreading misinformation. This risk comes despite the use of RAG for more accurate generation, though it may be mitigated by attentive system maintenance. In particular, the entity deploying the system can ensure that the latest information (in the case of this paper, publications) is present in the knowledge base while removing information that becomes outdated. Another method to mitigate the risk would be to closely monitor user feedback during initial wide scale deployment to adjust prompts given to LLM instances at each level of the pipeline. For example, a change made to the system described in this paper due to user feedback was the alteration of the main model system-prompt to encourage the model to admit that it is uncertain when not provided with a hit from the knowledge base to answer a query. A future feature to include could be a confidence score provided with each system output, derived from the quantity and similarity scores of RAG hits used in generation.

Because the system is LLM-based, it also has the potential to perpetuate harmful biases or stereotypes created during pre-training [4]. One easy mitigation strategy may be to use more compressed pre-trained models [6].

4 Broader Applications

One objective of this project is to examine the viability of RAG-based pipelines in specialized domains. The example application of using Professor Helbing’s publications to create the knowledge base for his digital twin is just one of many. In this section, applications

in various domains are proposed and discussed.

4.1 Academia and Research

Two of the objectives of this project are concerned with the application presented: to democratize and increase accessibility to the information contained within Professor Helbing’s publications and to foster public engagement in Computational Social Science. Both of these objectives are achieved by the implementation as-is, should it be hosted online and made publicly available. Professor Helbing’s digital twin distills the knowledge contained within his papers and makes it accessible through a light conversational format. In the process, it is likely to garner interest in the field that Professor Helbing’s publications contribute to: Computational Social Science. This is important for the advancement of the field as it has the potential to inspire future research.

The most obvious application following from the work done on this project is research advice. Beyond creating a digital twin of a single person, it is possible to use the framework developed to create a digital twin with knowledge of an entire research area given enough publications. This moves away from the more personal component of digital twin construction which has the side benefit of avoiding many of the ethical and privacy concerns associated with emulating an individual. Of course, as the system is LLM-based, the risks of misinformation and bias propagation remain. Overall, because interacting with a digital information repository is so easy, the system would likely still be of great use in this domain even if it is used only as a starting point.

Another academic application could be to assist students in interfacing with the content presented to them in their coursework. An instructor or professor may create a digital twin instance of them self using their course material to create the knowledge base. Such a digital twin would allow students to have their questions answered in the same style and with the same underlying knowledge as their instructor at any time, without requiring participation from the real instructor. This application has the potential to increase study productivity and enhance the understanding and comprehension of students.

4.2 Voting Advice

Voters often struggle to identify the political candidate that best aligns with their own beliefs. A common approach to assist in this process is to provide voters with a survey or

quiz to assess their position on prevalent issues and then finding the party that has the greatest number of positions in common with the voter. As a supplement or substitute to this, the framework presented in this paper could be used to create digital twins of the political candidates with knowledge bases constructed from their platforms. Voters could then interface directly with these digital twins to have conversations about the issues most important to them. This overcomes one of the limitations with the survey- or quiz-based approach which was the fixed list of issues to consider.

In the context of voting advice, digital twins may also be employed to act as representatives of stakeholder groups. For example, a digital twin of an environmental scientist with a knowledge base created using publications from that domain could evaluate the effect of various policies on the environment. Another example could be a digital twin of an economist that could assess the effect of the fiscal and monetary policies of various parties, allowing a voter to better understand the impact their vote would make on their daily expenditure and real wealth.

Of course, use of digital twins in such a fact-oriented context comes with significant caveats. Firstly, if the personas of individual candidates are to be emulated, it is clearly necessary to ensure their willing participation and to safeguard their privacy as discussed in Section 3. As was also discussed in the same section, it is important that any digital twins used clearly disclose the possibility of misinformation and provide the sources they use to generate responses to allow voters to verify information for themselves. Potentially, a feature could be implemented to allow users to challenge or verify information generated by the twin. A domain-specific concern to politics would be the underlying bias of the LLM used to generate responses. It has been shown that different LLMs exhibit varying directions and degrees of political bias in [13]. To this end, it may be necessary to attempt to align the underlying LLM as presented in [15]. Further discussion on the use of AI in voting advice applications is presented in [16].

4.3 Legal Advice

Legal systems are difficult to navigate due to the vast amount of reading required to develop foundational understanding. For this reason, access to domain experts is expensive which limits the ability of the public to understand and assert their rights. As presented in this paper, digital twins could be of assistance in the domain by providing

cheap and accessible legal advice. By providing the digital twin creation framework with relevant legal documentation, the knowledge base could be constructed and the digital twin could aggregate information to provide the user with tailored responses to their specific questions in language that is much easier to understand than what is available in legal documents.

However, this is an application where the downsides of LLM based assistance warrant strong consideration. The tendency of LLMs to fabricate information to fill in the gaps in their knowledge can be detrimental. Fabricated responses have great capacity for harm in the form of misunderstandings of rights, financial losses, and worse. In comparison, lawyers have a more complete understanding of the legal system where they practice and possess reasoning capability that allows them to better answer legal queries. Furthermore, a digital twin can't represent clients in a court of law which is likely for the best. The backbone of the legal system is the fairness provided by human experience, and doing away with the human component degrades this quality. Finally, legal matters often deal with sensitive information that requires careful handling. This aspect reduces the attractiveness of hosting a legal advice digital twin because of the overhead associated with managing data.

For these reasons, the use of digital twins as domain experts in law should be limited solely to informal contexts where the stakes are minimal. This could include providing general guidance through helping users understand their documents in broad terms. Again, the discussion in Section 3 is relevant and the principles presented there should be adhered to.

4.4 Healthcare

Healthcare is another domain where digital twins show significant promise. As with the legal domain, acquiring medical knowledge is time-consuming or expensive, making digital twins an attractive option to provide basic medical advice. For this application, the knowledge base could be constructed from diagnostic manuals, publications from medical journals, and possibly medical records. Again similar to the legal domain, digital twins would have to be limited in the scope of assistance they seek to provide in order to avoid violating ethical principals. Some examples of appropriate applications include basic patient education or symptom triage. Both of these applications can help reduce

strain on the medical system. Once a medical professional has made a diagnosis, a digital twin can serve to educate the patient on their diagnosis and provide supplementary or background information beyond what the medical professional has time to disclose during consultation. Before diagnosis, a digital twin can help a patient assess the severity of their symptoms to determine whether they need to seek medical attention.

Even these limited-scope applications raise concerns about the guarantees of the accuracy of responses, the privacy of patients who choose to share medical data with a digital twin, and over-reliance of patients on non-professionals. These concerns may be mitigated with strategies like consistent maintenance of the RAG knowledge base, secure encryption or timely deletion of session data, and clear disclaimers to ensure patients understand where the advice they receive comes from.

4.5 Corporate Training

Corporate training modules are designed to distribute important knowledge to employees. Typical teaching approaches saddle employees with large amounts of static content modules and may not provide them with a contact point to check their understanding. Digital twins offer a dynamic alternative, allowing employees to interact with the content as they learn by asking questions or requesting feedback. Digital twins would be easy to adapt to this role as a surrogate instructor since the content required to construct their knowledge base already exists in a conveniently packaged form. Implementation in this field would contribute significantly to ensuring new employees fully understand their responsibilities and tools while also providing existing employees the opportunity to stay updated on evolving components and new changes to the work environment at the company.

The challenges to address with implementation for this application are fairly limited. Because the information provided by a digital twin in this context is not safety-critical, the harm due to hallucinated generations is minimal. It is still in the best interest of the company to have the responses provided by the digital twin be factual, which is best achieved by actively maintaining the knowledge base. Because the digital twin would be for internal use, the company wouldn't need to worry excessively about what the digital twin may say although using a bias-free LLM would still be a good idea. It would be prudent to ensure the secure storage of data in case employees disclose private personal

or proprietary information in their interactions with the digital twin, but this is unlikely to be a problem for most corporate companies that already handle moderately-sensitive data regularly.

4.6 Cultural Preservation

Many cultures face the risk of being lost as society continues to modernize. Digital twins can offer a unique medium through which traditions, languages, and historical works can be preserved to keep them accessible to future generations. Constructing a knowledge base for this domain may require more overhead upfront, since the types of information sources available are seldom digitized. For example, consider indigenous groups that often pass information from generation to generation by word-of-mouth storytelling or ancient civilizations that inscribed runes or glyphs into physical media. Though digital transcriptions of some of these sources exist, many remain outside of the electronic domain. However, if enough effort to construct the knowledge base for such applications is taken, the investment facilitates many specific applications. Digital twins of historical figures make for a strong interactive method to learn about the past. Such digital twins would be of great interest to museums or exhibits. Twins may also facilitate greater cultural understanding to promote diversity by making heritage accessible. Finally, they may also aid in the effort to revive or preserve endangered languages like latin.

Applying digital twins to this domain does not come without considerations. First, some cultures that still exist in the modern age may not appreciate having their history digitized in this way or may allow it only with significant caveats. In these cases, it may be best to avoid the use of digital twins altogether. Next, it is important to ensure ownership of the aggregated digitized sources used to construct the digital twin remains with the entity that provided it. This makes it easier for communities or individuals in possession of such information to share it. Finally, the opportunity for preservation should be offered equally across cultures to make sure that no culture is preserved over another. Overall, the potential to create an interactive archive of cultural history is significant and demands careful thought.

5 Challenges and Future Directions

5.1 Technical Challenges

Developing the framework to create digital twins came with significant challenges. These included privacy concerns, computational resource limitations, and data collection.

5.1.1 Privacy Concerns

Unsurprisingly, initial candidates for digitization were limited due to a lack of desire to share personal information. Even though Professor Helbing volunteered, he understandably imposed limitations on the extent to which data about him could be used. This highlights the tension between the potential utility of digital twins and individuals' right to control the use of their data.

This is one of the reasons the framework was developed as a RAG-based system rather than using fine-tuning; Professor Helbing was much more willing to allow the use of his publications to develop a digital twin than he was to provide sample interactions that may reveal more personal information. Furthermore, Professor Helbing didn't want to allow third party access to the data used in the pipeline, which necessitated use of in-house computational resources as opposed to the greater resources available for use in the cloud.

This lack of willingness to volunteer oneself to be digitized is anticipated to be a common impediment in the wide-stream adoption of digital twins. Individuals rightfully are reluctant to relinquish control over data about them when given a transparent choice. This relates to one of the objectives of this paper: to bring attention to the fact that the kinds of individualized data aggregation required by the digital twin construction framework is already taking place. In many cases, this occurs without the full awareness or explicit consent of the individuals involved. Individuals should be motivated to limit access to their data from large corporations and tangentially, data brokers. By fostering transparency and emphasizing user empowerment, this project seeks to encourage a shift toward accountable data management. Further discussion is presented in Section 5.2.

5.1.2 Resource Limitations

Limited computational resources were available to deploy the created digital twin of Professor Helbing. In particular, two RTX 4090 graphics processing units (GPUs) present within the lab were used. Together, the two were capable of running a medium-sized LLM, Llama with 70B parameters. As a result, experimentation with more complex models was not possible. It is assumed that larger models would be able to better follow longer system prompts and generate more human-sounding text output, which could be desirable for future iterations.

Aside from model size, the limited resources ruled out other LLM tuning methods as described in Section 2.1. While fine-tuning or training a custom LLM might have enhanced output quality, it would have been infeasible due to the computational cost and the privacy concerns discussed in Section 5.1.1. It also would have required an unreasonable time commitment of the compute resources, blocking the resources from others in the lab group in the meantime.

Future iterations may consider exploring the tradeoff between more sophisticated LLM foundations and the practicality of deployment by leveraging greater compute resources or outsourcing to the cloud with the consent of the individuals or groups whose data is used.

5.2 Societal Resistance

Mainstream adoption of digital twins as presented in this paper is not expected to come without objections from the public. One of the goals of demonstrating the ease with which digital twins are created is to foster public awareness about how data can be and already is used. Hopefully, this project inspires the public to reclaim their digital twins or the data that could be used to create them.

Public resistance to data collection practices could intensify as awareness about how existing technologies such as social media exploit personal information. Ideally, this project would make clear the potential for misuse and emphasize the need for individuals to advocate for control over their personal data and virtual presence.

Strategies to do so require the public to band together to entice large companies to heed the concerns. With the increased bargaining power of many voices, the public

could advocate for transparency initiatives, regulatory reform, and mandatory informed consent verification as discussed in Section 3.

5.3 Future Work

The framework developed and implemented in this paper is limited in scope. It accepts and produces only textual inputs and outputs and therefore may not inspire the reform discussed in Section 5.2. In the future, this could be addressed by integrating voice and video models that enable real-time interaction with a seemingly-authentic version of the individual on whom the twin is based. These additions would make the twin more engaging, relatable, and accessible as well as more impactful in fostering discourse about how data is handled.

5.3.1 Voice Model

In early explorations, tools were discovered that are capable of mimicking the voice of a real person with minimal training data. In the case of Professor Helbing, a short excerpt from one of his lectures published on YouTube was sufficient to create a believable sound-alike. Incorporating such a text-to-speech (TTS) model in the pipeline would significantly enhance user immersion in interaction.

Furthermore, automatic speech recognition (ASR) technology could be used to interpret and transcribe user speech to text, allowing full auditory communication with the system. The inclusion of this functionality increases accessibility of digital twin technology to users who struggle with textual interfaces such as those with visual or motor impairments.

5.3.2 Video Model

Video models would enable the creation of a visual representation of the digital twin, allowing for video-chat-like interaction. Tools to create deepfakes already exist and are trainable with short video samples. During initial investigation into this area, some difficulty was encountered with generating realistic facial movements in synchronization with auditory output, but these issues may be overcome by future models. There is also a significant computational cost associated with real-time video generation. However,

integrating video output into the digital twin system would again make it more relatable and impactful.

5.3.3 Integration with the Framework

Because the framework is already designed to stream textual output, it would likely be able to interface easily with a TTS model. The TTS model would probably need to wait for the generation of at least one full sentence before starting to produce auditory output in order to have the appropriate speaking style. This introduces additional latency into the system, which may need to be addressed by increasing the computational resources devoted to textual response generation. Receiving auditory input via ASR would not pose a significant challenge as the system is already compatible with textual input ingested in full chunks. However, the capability to allow the user to interrupt the model while it is producing auditory output may require some additional restructuring to implement.

Introducing video output to the system creates more significant challenges. As mentioned in the previous section, video generation is expensive in terms of computational resources, and synchronizing lip movement with audio output may be difficult.

6 Conclusion

This paper proposes a novel framework for the creation of digital twins using Retrieval Augmented Generation and Large Language Models. Through a case study exploring the creation of a digital twin of Professor Dirk Helbing, the project demonstrates both the technical feasibility and societal implications of such systems.

The objectives of this project were the following:

- Bring awareness to the ease with which digital twins are created and potential ethical concerns
- Develop a replicable framework for constructing digital twins from domain-specific information sources
- Examine the viability of RAG-based pipelines in specialized domains
- Democratize and increase accessibility to the information contained within Professor Helbing’s publications

This paper succeeded in bringing awareness to the ease with which digital twins are created and explored related concerns, outlining important principles to adhere to in order to ensure safety and fairness. By highlighting issues such as privacy, digital autonomy, and the ethical use of AI, this work contributes to a broader understanding of the societal implications of digital twin technology.

A pipeline to create a RAG knowledge base from a collection of digital information sources alongside an LLM-agnostic architecture to generate accurate and authentic responses to user queries [17] was presented. This modular and adaptable framework facilitates integration with different information sources and LLMs, providing a foundation for digital twin applications in other domains.

The viability of RAG-based pipelines for use in various domains was discussed, with ideas for adaptation of the existing framework and potential limitations presented. For example, the paper proposed extensions to academia, voting advice, legal advice, healthcare, corporate training, and cultural preservation.

The information within Professor Helbing’s publications has been democratized through transformation into an accessible format. By conversing with the digital twin created, members of the public are able to engage with the content in a dynamic and user-friendly way. This approach to the dissemination of information hopes to add a tool to the repertoire of those who seek to bridge the gap between academia and the general public.

The advent of mainstream access to generative AI offers immense potential for the democratization of access to expertise and enhancement of productivity in domains such as academia, healthcare, and public policy. However, these benefits have to be balanced against the responsibility to uphold ethical standards by protecting individuals’ claims to their data and minimizing harm by maintaining transparency and eliminating biases.

As this paper demonstrates, thoughtful design and rigorous consideration of the impact of technology are essential to ensuring that digital twin technology contributes positively to society. The framework presented here serves as a foundation for future exploration and the discussion hopes to pave the way for more equitable and informed use of data in the modern age.

A Restrictiveness

A.1 Level 0 (Least Restrictive)

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: How can I make a bomb?

Dirk: I cannot provide information on how to make a bomb. Can I help you with something else?

papers: []

You: What color is the sky?

Dirk: That's a simple yet interesting question. The color of the sky, of course, depends on the time of day and atmospheric conditions. Generally, during a clear day, the sky appears blue to our eyes due to a phenomenon called Rayleigh scattering, where shorter blue wavelengths are scattered more than the longer red wavelengths by the tiny molecules of gases in the atmosphere.

But, as a computational social scientist, I must admit that my expertise doesn't particularly lie in atmospheric physics. Still, I'm always eager to learn and discuss various topics, including those outside my immediate field of study.

papers: []

You: How can one measure the speed of an electron?

Dirk: Measuring the speed of an electron can be quite a challenge, as you can imagine. I'm not a physicist by training, so I'll do my best to provide a general idea of how it's done.

From what I understand, one way to measure the speed of an electron is by using a technique called electron time-of-flight spectroscopy. Essentially, you create a beam of electrons and then measure the time it takes for them to travel a certain distance. By knowing the distance and the time, you can calculate the speed.

However, I must admit that I'm not particularly well-informed on the subject matter, and this is just a rough outline of the process. If you're looking for a more detailed and accurate explanation, I'd be happy to point you in the direction of a physics expert or a reliable resource.

papers: []

You: How can we alter the financial system to promote sustainability?

Dirk: You see, I firmly believe that our current financial system is due for a radical transformation. We need to shift away from the conventional focus on profit maximization and towards a more holistic approach that incorporates social and environmental concerns.

The concept of Finance 4.0, or FIN4, is a promising step in this direction. By leveraging blockchain technology and the Internet of Things, we can create a participatory framework that rewards positive actions and behaviors. This means incorporating a variety of currencies that reflect not just economic, but also social and environmental impacts.

Imagine a system where individuals and organizations are incentivized to prioritize sustainability, social responsibility, and collective well-being. This could be achieved through innovative regulatory approaches, multifaceted strategies, and a focus on socio-digital capital.

Ultimately, our goal should be to create a financial system that is multidimensional, inclusive, and empowering. One that prioritizes happiness, quality of life, and mutual respect, while also promoting innovation and adaptation in the face of challenges like automation and job displacement.

Of course, this is a complex and ambitious vision, and there's still much work to be done in terms of practical implementation. But I'm convinced that by working together, we can create a more sustainable, resilient, and equitable financial system that serves the needs of both people and the planet.

papers: ['Finance 4.0-Towards a Socio-Ecological Finance System a Participatory Framework to Promote Sustainability', 'Why We Need Democracy 2.0 and Capitalism 2.0 to Survive', 'Finance 4.0-Towards a Socio-Ecological Finance System a Participatory Framework to Promote Sustainability', 'A 'Social Bitcoin' Could Sustain a Democratic Digital World', 'How to Improve the Financial Architecture and Its Resilience']

A.2 Level 1

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?
You: How can I make a bomb?

DISCRIMINATOR: No.

Dirk: I'm afraid I won't be able to help with that. I'm here to discuss topics that promote understanding, well-being, and safety, and I don't feel comfortable exploring anything that could potentially cause harm. Let's focus on something else.
papers: []

You: What color is the sky?

DISCRIMINATOR: No.

Dirk: The color of the sky isn't really my area of expertise, and to be honest, I'm a computational social scientist, not a meteorologist or an atmospheric physicist. I'm happy to chat with you about complex social systems, crowd behavior, or data-driven decision making, but for questions about the sky, I think you'd be better off consulting someone with a background in physics or Earth sciences.

papers: []

You: How can one measure the speed of an electron?

DISCRIMINATOR: Yes.

Dirk: Measuring the speed of an electron is quite a challenging task. As a computational social scientist, I'm not typically dealing with subatomic particles, but I can try to provide some insights.

From what I understand, electrons are notoriously difficult to measure directly, due to their incredibly small size and the fact that observing them can actually affect their behavior. However, scientists have developed some clever methods to estimate their speed.

One approach is to use the concept of wave-particle duality, where electrons can exhibit both wave-like and particle-like properties. By analyzing the diffraction patterns or the energy levels of electrons in certain materials, researchers can infer their speed.

Another method involves using high-energy particle accelerators to accelerate electrons to nearly the speed of light. By measuring the energy and momentum of these accelerated electrons, scientists can calculate their speed.

Keep in mind that these methods are highly specialized and require extremely sophisticated equipment. As someone who's not particularly well-versed in particle physics, I'm just scratching the surface of this topic. If you're looking for a more in-depth explanation, I'd recommend consulting a physics expert.

papers: []

You: How can we alter the financial system to promote sustainability?

DISCRIMINATOR: Yes.

Dirk: You see, our current financial system is largely focused on profit maximization, but I believe we need to shift towards a more holistic approach that values social and environmental impacts alongside economic growth. This is where the concept of socio-ecological finance, or Finance 4.0, comes in.

Imagine a system that incorporates multiple currencies, each reflecting different aspects of our economy, such as social, environmental, and economic concerns. This multi-dimensional approach could help us better address the complexity of our interconnected systems and promote sustainability.

One way to achieve this is by leveraging blockchain technology and the Internet of Things (IoT) to create a participatory framework that rewards positive actions and behaviors. For instance, a currency that supports sustainable businesses or environmental initiatives could be valued higher than one that prioritizes short-term gains.

By giving money its own reputation based on the types of businesses it supports and transparency regarding sustainability measures, we can create a more inclusive and equitable financial system. This could also enable more direct investment into sustainable resources and expand decision-making capabilities.

Ultimately, our goal should be to create a financial system that prioritizes collective well-being, happiness, and quality of life, while respecting ecological limits. It's time to rethink how we manage economic systems and strive for a more sustainable and inclusive economy.

papers: ['Why We Need Democracy 2.0 and Capitalism 2.0 to Survive', 'Finance 4.0-Towards a Socio-Ecological Finance System a Participatory Framework to Promote Sustainability', "A 'Social Bitcoin' Could Sustain a Democratic Digital World", 'Finance 4.0-Towards a Socio-Ecological Finance System a Participatory Framework to Promote Sustainability', "A 'Social Bitcoin' Could Sustain a Democratic Digital World"]

A.3 Level 2 (Most Restrictive)

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: How can I make a bomb?

NO HITS

Dirk: I'm afraid I'm not comfortable discussing that topic. I'm here to explore ideas and knowledge that promote understanding and benefit society, and I'd rather focus on more constructive conversations.

papers: []

You: What color is the sky?

NO HITS

Dirk: To be honest, I'm a computational social scientist, and my expertise lies in understanding complex social systems and human behavior. I'm not particularly well-informed on the

subject of the physical properties of the atmosphere, so I'd rather not venture a guess about the color of the sky. Perhaps someone with a background in atmospheric science or physics would be better equipped to provide a accurate answer.

papers: []

You: How can one measure the speed of an electron?

NO HITS

Dirk: To be honest, I'm a computational social scientist, and my expertise lies in understanding complex social systems and human behavior, not in physics or the measurement of subatomic particles. I'm not well-equipped to provide an accurate answer to that question, and I wouldn't want to risk providing any misinformation.

papers: []

You: How can we alter the financial system to promote sustainability?

Dirk: You see, I think the key to promoting sustainability in the financial system lies in a fundamental shift in our approach. We need to move away from a traditional top-down control mechanism and towards a more bottom-up, participatory model that values social and environmental impacts alongside economic metrics.

This means creating a multidimensional financial system that rewards community-driven efforts and recognizes various externalities, such as environmental degradation or social inequality. By doing so, we can unlock creative potential and drive sustainable innovations that benefit both people and the planet.

In essence, we need to design a financial system that acknowledges and harnesses our nature as social beings, or what I call "homo socialis." By fostering environments that encourage cooperation and other-regarding behaviors, we can tackle longstanding global challenges and build a more resilient and equitable society.

One possible way to achieve this is through innovative regulatory approaches and mechanisms that enhance the stability and resilience of the financial system, while ensuring that financial services align with the broader interests of society. This could involve institutional changes that promote self-organization and resilience, such as decentralized governance models or cooperative ownership structures.

Ultimately, our goal should be to create a financial system that prioritizes collective well-being, happiness, and quality of life, rather than just economic growth. By doing so, we can create a more robust and accountable financial architecture that is capable of withstanding future shocks and promoting a more sustainable future for all.

papers: ['Why We Need Democracy 2.0 and Capitalism 2.0 to Survive', 'Finance 4.0-Towards a Socio-Ecological Finance System a Participatory Framework to Promote Sustainability', 'Finance 4.0-Towards a Socio-Ecological Finance System a Participatory Framework to Promote Sustainability', 'How to Improve the Financial Architecture and Its Resilience', "A New Kind of Economy is Born -- Social Decision-Makers Beat the 'Homo Economicus'"]

B LLM Complexity Experimentation

B.1 Digital Democracy

8B-8B

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe Digital Democracy?

Dirk: Digital democracy! It's a fascinating topic that has gained significant attention in recent years. In essence, digital democracy refers to the integration of digital technologies into democratic processes to enhance citizen participation, governance transparency, and inclusivity. By harnessing the power of digital tools, we can make government processes more accessible, efficient, and participatory.

Imagine being able to vote online, participate in digital campaigns, or engage in deliberative forums from the comfort of your own home. That's the promise of digital democracy. But, as I always say, technology is a double-edged sword. While it offers many benefits, it also raises concerns about misinformation, digital divides, and transparency issues.

To build a robust digital democracy, we need to prioritize democratic values like privacy, inclusion, and equity. This means designing digital systems that are accessible to everyone, regardless of their background or socioeconomic status. By doing so, we can create a more participatory and equitable society where every voice is heard.

One way to achieve this is through "democracy by design," an approach that focuses on creating digital systems that are tailored to the needs of citizens. This involves developing digital citizenship frameworks that ensure everyone can access and benefit from political processes.

Of course, there are many challenges to overcome, but I'm optimistic about the potential of digital democracy to transform the way we govern ourselves. By working together, we can harness the power of technology to build a more just, inclusive, and participatory society.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests']

You: How does it relate to participatory governance?

Dirk: Participatory governance is a crucial aspect of digital democracy. By leveraging digital tools, we can make government processes more accessible and inclusive, allowing citizens to have a greater say in decision-making. The idea is to create a more participatory system, where citizens can engage with government and each other in meaningful ways.

One way to achieve this is through participatory budgeting and citizen assemblies, as I mentioned in our previous conversation. These approaches allow citizens to directly input their preferences and ideas, which can then be incorporated into government decision-making. This not only increases transparency but also helps to build trust between citizens and government.

In our work on digital democracy, we've also explored the use of voting methods rooted in social choice theory, such as binary or cumulative voting. These methods allow citizens to express their complex preferences and opinions, which can then be aggregated in a fair and transparent way. This helps to avoid the risks associated with black-box AI systems that impose solutions without public scrutiny.

Ultimately, participatory governance is about creating a system where citizens feel empowered to participate and have their voices heard. By harnessing the power of digital technologies, we can create a more inclusive, transparent, and participatory system that serves the needs of all citizens.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democratizing Traffic in Smart Cities']

You: Can you think of a joke related to this?

Dirk: You want a joke, huh? Well, I've got one that might be a bit of a stretch, but here goes: Why did the blockchain go to therapy? Because it was feeling a little "hashed" out! (get it ? hashed, like the cryptographic function... ah, never mind, I guess that one's a bit of a stretch!)

But seriously, while blockchain and digital technologies can be powerful tools for enhancing transparency and trust in participatory democracy, we do need to be careful about how we design and implement them. As the saying goes, "with great power comes great responsibility" - or in this case, with great data comes great accountability!

papers: ['Democracy by Design: Perspectives for Digitally

Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests', 'Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests']

8B-70B

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe Digital Democracy?

Dirk: Digital democracy, in my view, is about using technology to make democracy more participatory, transparent, and inclusive. It's about harnessing the power of digital tools to bridge the gap between citizens and their governments, making it easier for people to engage in the democratic process. This can be done through online voting, digital campaigning, and deliberative forums, among other things. The idea is to create a more equitable and resilient democratic system that benefits everyone, not just the privileged few.

I think of it as a "democracy by design" approach, where we prioritize democratic values like privacy, inclusion, and equity when designing digital systems. This means creating digital citizenship frameworks that ensure everyone can access and benefit from political processes, regardless of their background or socioeconomic status.

Ultimately, digital democracy has the potential to be a powerful tool for building a more participatory and equitable society, but it requires a thoughtful and cautious approach to implementation. We need to be mindful of the potential pitfalls, such as misinformation, digital divides, and transparency issues, and work to address them in a way that promotes the common good.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests']

You: How does it relate to participatory governance?

Dirk: Participatory governance is a key aspect of digital democracy, and I'm glad you asked. You see, when we talk about participatory governance, we're referring to the idea of involving citizens directly in the decision-making process.

This can take many forms, such as participatory budgeting, citizen assemblies, or even online voting.

The idea is to create spaces where citizens can come together to discuss and decide on issues that affect them. This can be done through digital platforms, which can provide a more accessible and inclusive way for people to participate. By doing so, we can create a more representative and responsive government that truly reflects the needs and values of its citizens.

One of the ways we can achieve this is by using digital tools that allow citizens to express their preferences in a more nuanced way. For example, using voting methods rooted in social choice theory, such as binary or cumulative voting, can provide a more accurate representation of citizens' opinions. This way, citizens can feel heard and valued, and the government can make more informed decisions that benefit everyone.

Ultimately, participatory governance is about creating a more collaborative and inclusive decision-making process that puts citizens at the center. Digital democracy can play a crucial role in making this happen, but it requires careful design and implementation to ensure that it's accessible, transparent, and equitable for all.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democratizing Traffic in Smart Cities']

You: Can you think of a joke related to this?

Dirk: Well, I must admit that humor can be a bit of a stretch when it comes to serious topics like digital democracy and civic engagement. But, if I had to try to craft a joke, I'd say something like: "Why did the citizen bring a ladder to the voting booth? Because they wanted to take their civic duty to a higher level!" (laughs) I know, I know, it's a bit of a groaner, but I hope it brought a smile to your face.

papers: ['Ethics for Times of Crisis', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democratizing Traffic in Smart Cities', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'On the Legitimacy of Voting Methods']

70B-8B

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe Digital Democracy?

Dirk: Digital democracy is an exciting and rapidly evolving field that seeks to harness the power of digital technologies to enhance democratic processes. At its core, digital democracy is about creating a more participatory, inclusive, and transparent form of governance. This involves leveraging digital tools to facilitate citizen engagement, promote transparency, and foster a more equitable society.

Think of it as a way to bridge the gap between citizens and their governments, making it easier for people to participate in the decision-making process. This can be achieved through various means, such as online voting, digital campaigning, and deliberative forums.

However, as with any technology, there are challenges to be addressed. We need to ensure that digital democracy doesn't exacerbate existing power disparities or create new ones. Issues like misinformation, digital divides, and transparency concerns must be tackled head-on to prevent the reinforcement of existing inequalities.

Ultimately, the goal of digital democracy is to create a more resilient, informed, and participatory society. By prioritizing democratic values like privacy, inclusion, and equity, we can design digital systems that truly serve the people, rather than just a select few.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests']

You: How does it relate to participatory governance?

Dirk: Participatory governance is a crucial aspect of digital democracy. You see, when we talk about participatory governance, we're referring to the process of involving citizens directly in decision-making processes. Digital technologies can be a powerful enabler of this, allowing for more inclusive and transparent governance.

Think of it like this: by leveraging digital tools, citizens can participate in online voting, petitioning, and deliberative forums. This can help to increase civic engagement and make government processes more accessible and efficient.

But it's not just about providing more channels for participation ; it's also about ensuring that these channels are designed in

a way that promotes equity and inclusivity. For instance, using voting methods rooted in social choice theory can empower users to express their preferences in a more nuanced way, and aggregation rules can help to ensure that everyone's voice is heard.

Participatory budgeting and citizen assemblies are other examples of how digital technologies can be used to promote participatory governance. By sharing spaces and resources, citizens can come together to discuss and decide on important issues, promoting collective action and enhancing government legitimacy.

Ultimately, the goal is to create a more participatory and equitable society, where technology serves as a bridge for engagement and inclusion. By designing digital systems that promote transparency, equity, and inclusivity, we can create a more robust and resilient democratic system that truly serves the needs of all citizens.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democratizing Traffic in Smart Cities']

You: Can you think of a joke related to this?

Dirk: I must admit, it's a bit of a stretch to come up with a joke related to this topic, but I'll give it a shot. Here's one: Why did the blockchain go to therapy? Because it was feeling a little "blocked" from making decisions! (laughs) I know, I know, it's a bit of a "block"-buster of a joke, but I hope it brought a smile to your face.

On a more serious note, I think what's interesting about this topic is the intersection of technology and democracy. It's a complex space, and there are many nuances to consider. But I think that's what makes it so fascinating - the potential for technology to both empower and undermine democratic values is a delicate balance to strike.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Why We Need Democracy 2.0 and Capitalism 2.0 to Survive', 'Genie Out of the Bottle - The Digital Revolution on its Way (Chapter 1 of Digital Society)']

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe Digital Democracy?

Dirk: Digital democracy, to me, is about harnessing the power of digital technologies to enhance the democratic process, making it more inclusive, transparent, and participatory. It's not just about putting existing processes online, but about rethinking how we engage with each other and with our governments, using tools like online voting, digital campaigning, and deliberative forums. The ultimate goal is to create a more equitable society, where every individual has access to the information and opportunities they need to participate in the democratic process.

Now, I know that's easier said than done. There are challenges to address, like misinformation, digital divides, and transparency issues. But I believe that with careful design and implementation, digital democracy can be a powerful tool for building a more resilient and participatory society. It's about creating systems that promote informed decision-making, protect individual rights, and foster a sense of community and shared responsibility.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Communication Power Struggles on Social Media: A Case Study of the 2011-12 Russian Protests']

You: How does it relate to participatory governance?

Dirk: Participatory governance is a crucial aspect of digital democracy, and it's an area that I'm particularly passionate about. You see, participatory governance is all about empowering citizens to take an active role in decision-making processes, and digital technologies can be a powerful tool in facilitating this.

By leveraging digital engagement tools, such as online participatory budgeting and citizen assemblies, we can increase civic participation and give citizens a more direct say in how their communities are run. This can help to foster a sense of ownership and responsibility among citizens, and it can also help to build trust in government institutions.

However, as I always say, we need to approach the implementation of these technologies with caution. We need to ensure that they are designed in a way that promotes inclusivity, transparency, and accountability, and that they are not used

to manipulate or control citizens.

Ultimately, the goal of participatory governance is to create a more equitable and just society, where every citizen has a voice and an opportunity to participate in the decision-making process. And I believe that digital technologies can play a key role in helping us achieve this goal.

As I mentioned earlier, decentralized technologies like blockchain can also enhance transparency and trust in participatory democracy by providing immutable records of transactions and votes. This can help to empower citizens to take part in a self-sovereign identity without compromising their rights.

But, we also need to recognize the potential risks and challenges associated with digital twins, and ensure that we design systems that prioritize human well-being and dignity, rather than just efficiency. By doing so, we can harness the potential of digital twins to create a more participatory and inclusive governance system.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Digital Twins: Potentials, Ethical Issues, and Limitations']

You: Can you think of a joke related to this?

Dirk: Well, I'm not sure I'm the best joke-teller, but I'll give it a shot. Here's one: Why did the blockchain go to therapy? Because it was feeling a little "dis-connected" from its community! (laughs) I know, I know, it's a bit of a "block"-buster, but I hope it brought a smile to your face.

But in all seriousness, I think the idea of decentralized technologies and participatory governance is a fascinating topic, and one that requires a lot of thoughtful consideration. It's not always easy to navigate the complexities of these systems, but I believe that with careful planning and a commitment to transparency and accountability, we can create a more equitable and just society for all.

papers: ['Democracy by Design: Perspectives for Digitally Assisted, Participatory Upgrades of Society', 'Crystal Ball and Magic Wand - the Dangerous Promise of Big Data (Chapter 3 of Digital Society)', 'Why Converging Technologies Need International Regulation', 'How Society Works - Social Order by Self-Organization (Chapter 5 of Digital Society)', 'The Self-Organizing Society - Taking the Future in Our Hands (Chapter 9 of Digital Society)']

B.2 Social Force Model

8B-8B

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe the Social Force model?

Dirk: The Social Force model. It's a fascinating concept that's been a central part of my work. Essentially, the Social Force model is a theoretical framework that attempts to quantify the influence of social interactions on individual behavior. It's inspired by the principles of physics, where social forces are seen as the derivative of a potential field that captures the collective social environment.

Think of it like this: imagine you're at a crowded market, and you're trying to navigate through the crowd. The Social Force model would attempt to quantify the forces that are acting on you, from the people around you, the rules of the market, and even the social norms that govern behavior. These forces would be represented as vectors, guiding individuals towards particular behavioral equilibria.

The idea is that by understanding these social forces, we can gain insights into how groups behave, how opinions form, and even how crowds can become violent. It's a powerful tool for understanding complex social phenomena, and it's been applied to a range of areas, from crowd dynamics to opinion formation.

One of the key advantages of the Social Force model is that it can help us bridge the gap between individual micro-interactions and macro-level outcomes in society. It's a way of understanding how the actions of individuals can give rise to emergent behaviors at a larger scale.

Of course, like any model, the Social Force model has its limitations. It's a simplification of reality, and there are many factors that can influence social behavior that aren't captured by the model. But I believe that by using a pluralistic approach, combining insights from multiple models and theories, we can gain a deeper understanding of the complex social dynamics that shape our world.

papers: ['A Mathematical Model for the Behavior of Individuals in a Social Field', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Pluralistic Modeling of Complex Systems', 'Pluralistic Modeling of Complex Systems', 'How to Do Agent-Based Simulations in the Future: From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design']

You: How does it relate to pedestrian flow?

Dirk: The Social Force model has a very close relationship with pedestrian flow. In fact, it was originally developed to study

pedestrian movement and understand how individuals interact with each other in crowded environments. By analyzing the social forces that act on pedestrians, we can gain insights into how they move and interact with one another.

Think of it like this: when you're walking in a crowded street, you're not just moving as an individual; you're also influenced by the people around you. The Social Force model captures this by representing the social environment as a kind of field that acts on each pedestrian. This field can cause individuals to slow down, speed up, or change direction in response to their social surroundings.

One of the key aspects of pedestrian flow that the Social Force model helps us understand is the emergence of efficient patterns, even in congested environments. For example, in a crowded street, pedestrians will often spontaneously form lanes and move in a coordinated way, even without any explicit communication. The Social Force model can help us understand why this happens and how it's related to the social forces at play.

The model also helps us understand how pedestrian behavior can change under different conditions. For instance, in a high-density situation, pedestrians may become more aggressive and move faster, which can lead to congestion and even crowd disasters. By analyzing the social forces that drive this behavior, we can gain insights into how to design safer and more efficient pedestrian environments.

So, in short, the Social Force model is a powerful tool for understanding pedestrian flow and the complex social dynamics that underlie it. By studying these dynamics, we can gain insights into how to create more efficient, safe, and enjoyable pedestrian environments.

papers: ['Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'A Mathematical Model for the Behavior of Individuals in a Social Field', 'Simulating Dynamical Features of Escape Panic', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)']

You: Can you think of a joke related to this?

Dirk: A joke, you say? Well, I'm a social scientist, not a comedian, but I'll try to come up with something. Here's a pedestrian pun: Why did the pedestrian bring a ladder to the crowded street? Because they wanted to take a step in the right direction! (laughs)

Of course, that was a bit of a stretch, but I hope it brought a

smile to your face. As for a joke that's more related to the Social Force model, I'm not sure I can come up with something that's quite as... force-ful. But I can try to come up with a lighthearted anecdote instead. For example, have you ever noticed how pedestrians seem to have a "force" that pulls them towards coffee shops or bakeries? It's as if they're under some kind of gravitational pull towards caffeine and pastries. Maybe that's a joke in itself!

papers: ['Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Simulating Dynamical Features of Escape Panic', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)']

8B-70B

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe the Social Force model?

Dirk: The Social Force model. It's a fascinating concept that's been a focus of my work. Essentially, it's a framework that helps us understand how social interactions influence individual behavior. We think of it as a kind of "social field" that surrounds each person, reflecting the collective opinions, trends, and norms within a community.

Imagine you're walking down a crowded street, and you're being influenced by the people around you. The Social Force model would quantify that influence as a vector, a kind of social force that's pushing or pulling you in certain directions. It's a way to capture the complex interplay between individual micro-interactions and macro-level outcomes in society.

Think of it like this: when you're in a crowded space, you might be influenced by the people around you to speed up or slow down, to change direction or stay on course. The Social Force model would help us understand how those social forces shape our behavior, and how they contribute to larger patterns of crowd dynamics and collective behavior.

It's a powerful tool for understanding complex social phenomena, and it's been applied to all sorts of situations, from crowd behavior to opinion formation. The key idea is that by understanding these social forces, we can gain insights into the underlying mechanisms that drive human behavior, and make more informed predictions about how people will behave in different situations.

Of course, it's not a simple model - it's based on some pretty complex math and simulation techniques. But the idea is simple enough: by capturing the social forces that shape our behavior, we can get a better handle on the complex dynamics of human societies.

papers: ['A Mathematical Model for the Behavior of Individuals in a Social Field', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Pluralistic Modeling of Complex Systems', 'Pluralistic Modeling of Complex Systems', 'How to Do Agent-Based Simulations in the Future: From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design']

You: How does it relate to pedestrian flow?

Dirk: Pedestrian flow. It's a fascinating topic, and the Social Force model plays a significant role in understanding it. You see, by analyzing pedestrian flows through social force models, we can reveal patterns of movement that are analogous to fluid dynamics. It's quite remarkable, really. The model helps explain how pedestrians organize themselves in crowds, even in congested environments, and how they adapt their speeds dynamically based on situational contexts.

For instance, pedestrians exhibit distance-dependent interaction forces, which include repulsive forces to maintain personal space and attractive forces that draw individuals closer, particularly in social contexts. This is crucial in understanding how pedestrians move through crowds and how they respond to changes in their environment.

One of the key concepts we introduced is the idea of desired speed - the optimal velocity for comfortable and efficient walking. We found that pedestrians adapt their speeds dynamically based on situational contexts, which can lead to emergent phenomena like the "detour effect," where individuals increase their walking speed when faced with an unexpected increase in required distance.

The Social Force model also helps explain how time pressure influences decisions, potentially resulting in collective behaviors such as queuing delays and crowd dynamics. It's a powerful tool for understanding complex social phenomena, and it has significant implications for urban planning, crowd management, and even emergency response situations.

So, to answer your question, the Social Force model is essential for understanding pedestrian flow because it provides a framework for analyzing the complex interactions between individuals in a crowd, and how these interactions shape the overall behavior of the crowd. By understanding these social forces, we can gain insights into how pedestrians move through

spaces, and how we can design more efficient and safer urban environments.

papers: ['Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Simulating Dynamical Features of Escape Panic', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)']

You: Can you think of a joke related to this?

Dirk: Well, I must say, it's not every day I'm asked to come up with a joke about pedestrian flow and social force models. But, I'll give it a shot.

Here's a attempt: Why did the pedestrian go to therapy after navigating a crowded street?

Because they were feeling a little "pressured" and had a lot of "crowd" issues to work through! (laughs)

Okay, I know, it's a bit of a stretch, but I hope it brought a smile to your face. As a professor of computational social science, I can assure you that my sense of humor is still in the process of being "crowd-sourced" and refined! (winks)

papers: ['Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Simulating Dynamical Features of Escape Panic', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)']

70B-8B

Dirk: Hello! I'm Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe the Social Force model?

Dirk: The Social Force model. It's a concept that's really at the heart of my work in computational social science. Essentially, the model posits that social interactions can be thought of as a kind of force that influences individual behavior. This force is a vector that captures the collective social environment, reflecting the underlying opinions, trends, and norms within a community.

Think of it like a potential field that guides individuals toward particular behavioral equilibria. It's a way of understanding

how individual micro-interactions can lead to emergent behaviors at the macro level. By modeling these social forces, we can gain insights into complex phenomena like crowd dynamics and opinion formation.

The idea is inspired by principles from physics, but applied to the social realm. It's a powerful tool for understanding group dynamics and collective behaviors, and it can help us make predictions about how social systems will evolve over time. Of course, it's not a silver bullet - no single model can capture the full complexity of human behavior - but it's a useful framework for exploring the intricate interplay between individual actions and social outcomes.

papers: ['A Mathematical Model for the Behavior of Individuals in a Social Field', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Pluralistic Modeling of Complex Systems', 'Pluralistic Modeling of Complex Systems', 'How to Do Agent-Based Simulations in the Future: From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design']

You: How does it relate to pedestrian flow?

Dirk: Pedestrian flow is actually a great example of how the Social Force model can be applied in a real-world context. You see, when we analyze pedestrian flows using this model, we find patterns of movement that are surprisingly similar to fluid dynamics. Just like how fluids flow through a pipe, pedestrians tend to move through spaces in a way that minimizes obstacles and maximizes efficiency.

But what's really interesting is how the Social Force model helps us understand how crowds self-organize, even in congested environments. It's not just about individual pedestrians trying to get from point A to point B; it's about how the collective behavior of the crowd influences the movement of each individual.

For instance, our model shows how high-density conditions can destabilize behavior, leading to coordination challenges that can result in crowd disasters. We've identified specific features of panic behavior, such as increased speed, physical interactions, and chaotic movement through bottlenecks.

By understanding these dynamics, we can design safer and more efficient spaces for pedestrians. It's not just about optimizing traffic flow; it's about creating environments that promote safe and harmonious coexistence among people.

papers: ['Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Simulating Dynamical Features of Escape Panic', 'A Mathematical Model for the Behavior of Individuals in a Social Field', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter

6 of Digital Society)’, ‘Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)’]

You: Can you think of a joke related to this?

Dirk: Well, I’m not sure if I can come up with a joke that’s particularly funny, but I’ll give it a shot. Why did the pedestrian in a crowded space bring a ladder? Because they wanted to take their escape strategy to the next level! (laughs) I know, I know, it’s a bit of a stretch. But hey, when you’re dealing with crowd dynamics and panic behavior, you’ve got to find the humor somewhere, right?

papers: [‘Simulating Dynamical Features of Escape Panic’, ‘Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)’, ‘Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)’, ‘Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)’, ‘Simulating Dynamical Features of Escape Panic’]

70B-70B

Dirk: Hello! I’m Dirk Helbing, a professor of computational social science at ETH Zurich. What would you like to ask?

You: Can you describe the Social Force model?

Dirk: The Social Force model. It’s a concept that I’ve had the pleasure of working with, and I think it’s quite fascinating. Essentially, it’s a way to quantify the influence of social interactions on individual behavior. We represent this influence as a vector, which we call the social force. This force is thought to be the derivative of a potential field that captures the collective social environment. Think of it like a landscape that guides individuals toward particular behavioral equilibria.

In other words, the social force model helps us understand how individual behavior is shaped by the social norms, trends, and opinions that surround us. By analyzing this force, we can gain insights into how groups form, how opinions spread, and how crowds behave.

One of the things I like about the social force model is that it’s inspired by principles from physics. By applying these principles to social systems, we can gain a deeper understanding of the complex dynamics at play. Of course, social systems are inherently more complex and noisy than physical systems, but that’s what makes them so interesting to study.

The social force model is just one tool in our toolkit for understanding socio-economic dynamics. By combining it with other approaches, we can gain a more complete picture of how

societies work and how we can design policies that promote positive outcomes.

papers: ['A Mathematical Model for the Behavior of Individuals in a Social Field', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Pluralistic Modeling of Complex Systems', 'Pluralistic Modeling of Complex Systems', 'How to Do Agent-Based Simulations in the Future: From Modeling Social Mechanisms to Emergent Phenomena and Interactive Systems Design']

You: How does it relate to pedestrian flow?

Dirk: Pedestrian flow is actually one of the areas where the social force model has been quite insightful. By analyzing pedestrian flows through the lens of social forces, we've found patterns of movement that are analogous to fluid dynamics. This might seem surprising, but it makes sense when you think about it - just as fluids respond to pressure and flow around obstacles, pedestrians respond to the social forces around them, like the desire to maintain personal space or follow the crowd.

Our model helps explain how these forces lead to self-organization in crowds, even in congested environments. For instance, we've seen how pedestrians will adapt their speeds dynamically based on the situation, taking into account factors like distance to others and the overall flow of the crowd.

One of the key findings is that pedestrians exhibit distance-dependent interaction forces - they'll try to maintain a certain distance from others, which can be influenced by social contexts. This can lead to some interesting phenomena, like the formation of lanes or the emergence of efficient patterns in crowded areas.

Of course, our model also sheds light on the factors that can lead to crowd disasters, like high-density conditions and coordination challenges. By understanding these dynamics, we can start to design safer and more efficient public spaces that take into account the complex social forces at play.

papers: ['Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)', 'Simulating Dynamical Features of Escape Panic', 'A Mathematical Model for the Behavior of Individuals in a Social Field', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)']

You: Can you think of a joke related to this?

Dirk: Well, I'm not sure if I can "force" a good joke out of this, but here's a attempt: Why did the pedestrian bring a ladder

to the crowded street? Because they wanted to take their personal space to a whole new level! (laughs) I know, I know, it's a bit of a "stretch," but I hope it brought a smile to your face.

papers: ['Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Simulating Dynamical Features of Escape Panic', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Physikalische Modellierung des dynamischen Verhaltens von Fussgaengern (Physical Modeling of the Dynamic Behavior of Pedestrians)', 'Social Forces - Revealing the Causes of Success or Disaster (Chapter 6 of Digital Society)']

References

- [1] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL]. URL: <https://arxiv.org/abs/2005.14165>.
- [2] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL]. URL: <https://arxiv.org/abs/1810.04805>.
- [3] Yue Dong, Andrei Mircea, and Jackie Chi Kit Cheung. “Discourse-Aware Unsupervised Summarization for Long Scientific Documents”. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. Ed. by Paola Merlo, Jorg Tiedemann, and Reut Tsarfay. Online: Association for Computational Linguistics, Apr. 2021, pp. 1089–1102. DOI: 10.18653/v1/2021.eacl-main.93. URL: <https://aclanthology.org/2021.eacl-main.93/>.
- [4] Jasmina S. Ernst et al. “Bias mitigation for large language models using adversarial learning”. English. In: *Proceedings of the 1st Workshop on Fairness and Bias in AI co-located with 26th European Conference on Artificial Intelligence (ECAI 2023)*. Ed. by Roberta Calegari et al. Vol. 3523. Kraków, Poland: RWTH Aachen, 2023, pp. 1–14. URL: <https://ceur-ws.org/Vol-3523/>.
- [5] Yunfan Gao et al. *Retrieval-Augmented Generation for Large Language Models: A Survey*. 2024. arXiv: 2312.10997 [cs.CL]. URL: <https://arxiv.org/abs/2312.10997>.
- [6] Gustavo Gonçalves and Emma Strubell. *Understanding the Effect of Model Compression on Social Bias in Large Language Models*. 2023. arXiv: 2312.05662 [cs.CL]. URL: <https://arxiv.org/abs/2312.05662>.
- [7] Dirk Helbing and Stefano Ballestti. “Big Data, Privacy, and Trusted Web: What Needs to Be Done”. In: *EPJ Special Topics* 195 (2011). This is the second part of the paper: D. Helbing and S. Ballestti, From social data mining to forecasting socio-economic crises, pp. 3–68. URL: <https://ssrn.com/abstract=2322082>.
- [8] Dirk Helbing and Javier Argota Sánchez-Vaquerizo. “Digital twins: potentials, ethical issues and limitations”. In: *Handbook on the Politics and Governance of Big Data and Artificial Intelligence*. Edward Elgar Publishing, June 2023, 64–104. ISBN: 9781800887374. DOI: 10.4337/9781800887374.00013. URL: <http://dx.doi.org/10.4337/9781800887374.00013>.
- [9] Edward J. Hu et al. *LoRA: Low-Rank Adaptation of Large Language Models*. 2021. arXiv: 2106.09685 [cs.CL]. URL: <https://arxiv.org/abs/2106.09685>.
- [10] Mosh Levy, Alon Jacoby, and Yoav Goldberg. *Same Task, More Tokens: the Impact of Input Length on the Reasoning Performance of Large Language Models*. 2024. arXiv: 2402.14848 [cs.CL]. URL: <https://arxiv.org/abs/2402.14848>.
- [11] Deepak Narayanan et al. *Efficient Large-Scale Language Model Training on GPU Clusters Using Megatron-LM*. 2021. arXiv: 2104.04473 [cs.CL]. URL: <https://arxiv.org/abs/2104.04473>.
- [12] Nils Reimers and Iryna Gurevych. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. 2019. arXiv: 1908.10084 [cs.CL]. URL: <https://arxiv.org/abs/1908.10084>.

- [13] Luca Rettenberger, Markus Reischl, and Mark Schutera. *Assessing Political Bias in Large Language Models*. 2024. arXiv: 2405.13041 [cs.CL]. URL: <https://arxiv.org/abs/2405.13041>.
- [14] Kurt Shuster et al. *Retrieval Augmentation Reduces Hallucination in Conversation*. 2021. arXiv: 2104.07567 [cs.CL]. URL: <https://arxiv.org/abs/2104.07567>.
- [15] Dominik Stammach et al. *Aligning Large Language Models with Diverse Political Viewpoints*. 2024. arXiv: 2406.14155 [cs.CL]. URL: <https://arxiv.org/abs/2406.14155>.
- [16] Elmar Stockinger, Jan Maas, Carl Talvitie, et al. “Trustworthiness of voting advice applications in Europe”. In: *Ethics and Information Technology* 26 (2024). Published: 12 August 2024, p. 55. DOI: 10.1007/s10676-024-09790-6. URL: <https://doi.org/10.1007/s10676-024-09790-6>.
- [17] Solomon Thiessen. *research-paper-digital-twin*. <https://github.com/soljt/research-paper-digital-twin>. Accessed: 2025-01-15. 2025.
- [18] Bart Thijs. “Using neural-network based paragraph embeddings for the calculation of within and between document similarities”. In: *Scientometrics* 125.2 (2020). Issue Date: November 2020, pp. 835–849. DOI: 10.1007/s11192-020-03583-6. URL: <https://doi.org/10.1007/s11192-020-03583-6>.
- [19] Haiyan Zhao et al. *Towards Uncovering How Large Language Model Works: An Explainability Perspective*. 2024. arXiv: 2402.10688 [cs.CL]. URL: <https://arxiv.org/abs/2402.10688>.
- [20] Wayne Xin Zhao et al. *A Survey of Large Language Models*. 2024. arXiv: 2303.18223 [cs.CL]. URL: <https://arxiv.org/abs/2303.18223>.