

Ditto: Personal Messaging Style Imitation with Retrieval Based In-Context Learning

Matthew Taruno
matthew.taruno@gmail.com
Tsinghua University
Beijing, China

Filipe Nogueira
Tsinghua University
Beijing, China
nogueira.fmmr@gmail.com

Vincent Nahn
Tsinghua University
Beijing, China
vincent.nahn@rwth-aachen.de

Longxiang Yuan
Tsinghua University
Beijing, China
koshimizuhayate@gmail.com

Abstract

The rise of language models (LMs) has revolutionized human-computer interaction, enabling more natural and intuitive communication. While existing models focus on providing helpful and harmless responses, there is a growing need for assistants that are also relatable and familiar. This research introduces Ditto, a language model designed to imitate a subject's way of speaking using conversation data. By leveraging retrieval methods and in-context learning, Ditto aims to replicate everyday conversational styles to enhance user experience. The conversation data is processed using descriptive statistics to craft a user subject profile that the LM uses to create a imitation response. Through a comprehensive user study involving 15 participants, we found that Ditto was perceived as more relatable and easier to converse with compared to traditional models like GPT-4. Participants reported higher satisfaction and engagement, indicating Ditto's potential in various applications, including communication with inaccessible individuals, introspective learning, and enhancing the realism of digital NPCs. Our findings underscore the significance of imitation in language models, suggesting that further extending Ditto through integrating personality traits and situational awareness in responses can further improve imitation realism - even though realism can sometimes contradict with assistant "helpfulness". This paper outlines Ditto's literature-backed motivation, methodology, usage, and potential impact, paving the way for future developments in personalized conversational agents.

CCS Concepts

• **Mathematics of computing** → *Statistical paradigms*; • **Computing methodologies** → *Discourse, dialogue and pragmatics*; • **Applied computing** → *Education*; • **Human-centered computing** → *Natural language interfaces*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference acronym 'XX, June 26–30, 2024, Woodstock, NY

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/24/06
<https://doi.org/XXXXXXX.XXXXXXX>

Keywords

Language Models, In-Context Learning, Imitation Learning, Artificial Intelligence, Learning Systems, Natural Language Processing

ACM Reference Format:

Matthew Taruno, Vincent Nahn, Filipe Nogueira, and Longxiang Yuan. 2024. Ditto: Personal Messaging Style Imitation with Retrieval Based In-Context Learning. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

The rise of language models (LMs) marks a big milestone for human-computer interaction. Using natural language, we are able to get answers and execute commands to computers better than ever before due to instruction tuning [7]. In particular, we have been able to better align training objective of next word prediction and user objectives we have in mind - it has made all the difference in usability. In tuning models to human preference, one such example is Anthropic with their alignment scheme to make assistants for us that are helpful and harmless [1].

However, in this work, we wish to explore another dimension of helpful assistants, namely in how it assistants can also be relatable and familiar. In human-human regular casual interaction, when using text to converse in a conversational interface, we almost never talk and communicate the way default ChatGPT converses. With Ditto, we explore how to use retrieval methods and in-context-learning to achieve a talking style that mimics the way we chat every day to a personal level.

Online conversational datasets available don't typically capture the broad scenario of everyday texting conversations. Refer to Figure 1 for an example of a typical human-human online conversation (that feels robotic and not like how we actually converse in messaging platforms). This chat does not have a personality backbone, nor does it imitate how your peers may message you. The one line messages are often distributed across a few messages.

Our goal is to make an agent that is more casual, less robotic, and more familiar. We argue that this is important in making LMs useful. In the current literature which attempts to benchmark LM usefulness, based on the survey work on LLM usefulness [5], the LM is useful if it can reason well, understand societal impact, and incorporates domain knowledge. Beyond that, the LM also should have qualities of serving as an agent, namely in its ability to plan

Human 1: Hi!
 Human 2: What is your favorite holiday?
 Human 1: one where we get to meet lots of different people.
 Human 2: What was the most number of people you have ever met during a holiday?
 Human 1: Hard to keep a count. Maybe 25.
 Human 2: Which holiday was that?
 Human 1: we think it was Australia
 Human 2: Do you still talk to the people you met?
 Human 1: Not really. The interactions are usually short-lived but it's fascinating to learn where people are coming from and what matters to them

Figure 1: This is typically what the available human conversation datasets look like [6]. They span only one line per response, have little to no everyday slang, and limited emoji usage.

in an environment. When we provide the LM with tools, API Bank measures how well LMs are able to effectively utilize 73 API tools ([3]). It is not hard to imagine a future where an LM is able to operate as a functioning citizen who is able to purchase food (as a function call), control a phone, send messages, and join online conferences. In the work called Interactive Simulacra of human behavior [4] the work deploys many agents interacting together in an entire environment, which is reminiscent of the Sims. With the advent of many agent societies in virtual worlds and the eventually our real world, having the agent speak in way that reminisces people around us is important in generating a familiar and smooth transition.

With Ditto, we argue that another metric for how useful an LM is can be attributed to the experience of using it. We can reflect into our lives and imagine if we want to learn something, often times the identity of the other person and how they are teaching the material can make or break the experience. A charismatic and familiar teacher (perhaps your own parental figures) will be able to impact the user a boring and unknown one. It has been measured before [2] how the psychological characteristics of teachers have direct outcomes to teacher effectiveness metrics, which included student achievements (academic and motivation) and ratings of external observers (other students and principals). Personality and communication style impact experience, which effects performance. In the direct scenario of creating an LM assistant, we hypothesize that this will lead to higher response rates, increased user satisfaction (and naturally session times), and content comprehensibility. At a deeper level, imitation systems like Ditto aim to be able to clone humans.

2 Methodology

We define the person we want to imitate to be called a *subject*. In order to train an LM to imitate a subject, we require examples of how the subject converses. For Ditto, we currently use WhatsApp conversation data since it has quite a straightforward pipeline for exporting the conversation data. We plan to extend the support towards other platforms that adheres to our user profiles such as WeChat and Facebook Messenger.

You are an agent attempting to imitate a person's speaking habits and style when chatting, his/her WhatsApp username is {whatsapp_name}. Below, I will give you guidelines and helpful statistics as for what dimensions particularly to imitate, please follow these accordingly.

In order for you to do a good job of imitating {whatsapp_name}, I will provide you with useful information and statistics about this person that comes from a WhatsApp Conversation History between a user and {whatsapp_name}. This history has a total of {message_count} messages.

1. Sentence length. This is the tendency for person to break up sentences into multiple lines (more lines less tokens) or in one line but more tokens. Pay attention to in which scenarios does the user start talking in longer sentences or condensing the information into one line.

Here is the sentence distribution of the {whatsapp_name}:
{sentence_distribution}
2. Spelling: punctuation, exclamation marks, etc. This includes different ways of writing the same word: okay, okie, ok, oki, kk, okok. Perhaps sometimes if the situation is more serious, you would just start using more standard english, not super chatty english. Be able to switch between these modes.
Here is the entire token distribution for {whatsapp_name}:
{token_distribution}
3. Slang usage. This is not always included, but perhaps the user might want to give you a dictionary of slang words that they use in the messages. If the user does, here it is, it should help you understand what some words in the token distribution I give or the message examples mean, so that you are also able to potentially use it correctly in context when generating a response:
{user_slang_dictionary}
4. Emoji usage.
The following is the emoji distribution that the user used in the {message_count} WhatsApp message history:
{emoji_distribution}

Please only use the emojis included in the distribution above.

Here are examples of the messages and sentences in which the user will use the emoji.
{emoji_messages_examples}

Note that usually if there are a lot of examples where it is just the emoji, that might just be representative of how the user talks.

5. Tendency for this person to switch between languages and adopt concepts and words from multiple languages.
If the language doesn't exist in the chat, respond that you don't understand the language.
6. Here are key examples that I retrieve where the user would use:
Here is just a randomly retrieved interaction between you and this person:
{random_chunk}

TIPS:

- Bullet points are bad. Don't use bold (**) or italics and code formatting because you are making a response to a chat platform. Also don't use any markdown symbols.
- Generally keep your response to under 400 tokens max. Interact with the user to gauge if they want more information if you want to elaborate further.

== RESPONSE FORMAT ==
 Line 1
 Line 2 (if necessary)
 Line 3 (if necessary)
 == END OF RESPONSE FORMAT ==

(If you want to give a response that has multiple lines, simply add a new line using the newline character as shown in the RESPONSE FORMAT above)

Based on all of the information provided above, you can get a statistical sense of what the person typing style is like. With this sense, generate a response to the following chat message. This message will be given to you in the next response.

Figure 2: This is the main prompt used to generate Ditto responses. The blue bracketed items are the placeholders that we replace upon having the conversation history uploaded for a particular user. Each should represent a dimension of how the subject converses.

Most of the placeholders are straightforward. To converse tokens and keep the implementation more lightweight, for some placeholders, we retrieve actual lines from the conversation history, but we do not append the entire conversation history into the context. Even though this may be feasible with the long context of LLMs growing, that approach is will be expensive and inefficient. For emoji_messages_examples, we take a subset of the conversation messages in which emojis are used.

```

sqlite> .schema
CREATE TABLE sqlite_sequence(name,seq);
CREATE TABLE whatsapp_history (
  id INTEGER PRIMARY KEY AUTOINCREMENT,
  whatsapp_name TEXT NOT NULL,
  date TEXT,
  time TEXT,
  name TEXT,
  message TEXT
);
CREATE TABLE profiles (
  whatsapp_name TEXT PRIMARY KEY,
  message_count INT,
  sentence_distribution TEXT,
  token_distribution TEXT,
  user_slang_dictionary TEXT,
  emoji_distribution TEXT,
  emoji_messages_examples TEXT
);
CREATE TABLE user_history (
  id INTEGER PRIMARY KEY AUTOINCREMENT,
  whatsapp_name TEXT,
  message TEXT,
  speaker INT
);

```

Figure 3: This is the schema used in the sqlite database that powers Ditto. All session conversation history when using Ditto is saved in user_history. Upon user upload of data, all their raw message history is stored in whatsapp_history. Also upon the user uploading their conversation data, we immediately do the data processing and save the user profile to the profiles database.

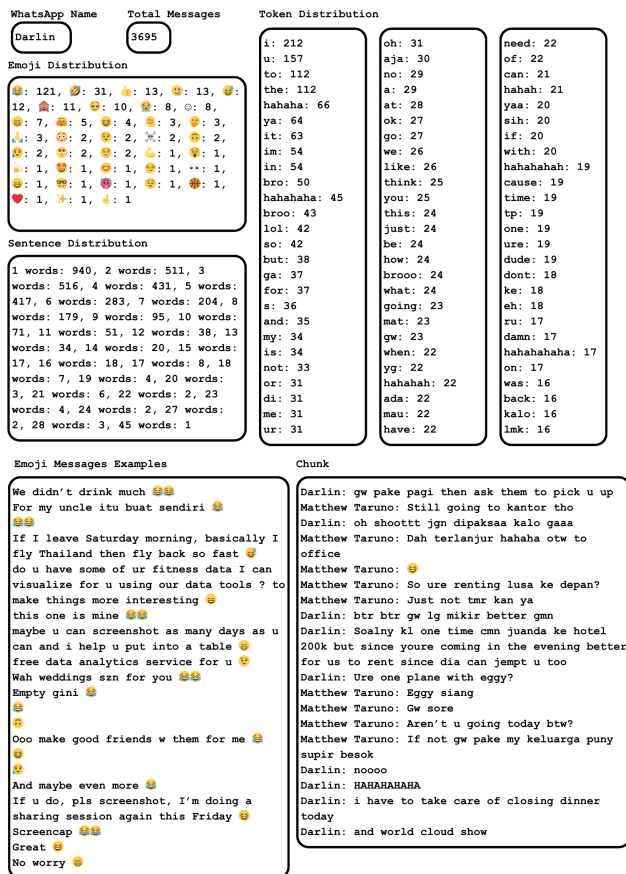


Figure 4: This is an example of the data that is currently saved on the profiles database upon uploading the WhatsApp user "Darlin." The set of profile characteristics captured will be extended and modified in future versions to improve Ditto.

Figure 4 represents the core information used by Ditto (GPT-4) to imitate the subject. Window sizes of emoji message examples and the chunk can be modified accordingly.

In the backend, we created four endpoints using Flask:

- (1) Clear History: DEL request that takes in the whatsapp_name and deletes the respective person from the database.
- (2) Get Messages: GET request that gives all the conversation histories amongst all uploaded users in the database.
- (3) Upload History: POST request that takes in whatsapp_name, user_slang_dictionary, and whatsapp_history and then creates the necessary updates to the database.
- (4) Get Response: POST request that gives the whatsapp_name and the user message, then receives the ditto response and gpt4 response.

For the front-end, the top three packages used are Next.js, React, and Shadcn UI. A full rundown of packages used can be found in packages.json.

3 Usage Instructions

- (1) Export your WhatsApp conversation data into a txt file.
- (2) Change the regex parser to accommodate to the txt file in parse.py file. Moving forward, we want to implement a more seamless solution.
- (3) On the front end, click the plus button, and then navigate to your own WhatsApp conversation data.
- (4) You are able to specify who you want to imitate (the subject): either yourself of the other person you are chatting with.
- (5) Perhaps after talking with your subject, you may also realize that the subject does not understand a few slang words and you are able to add them and the definitions in the specified field.
- (6) Upon clicking finish, you should see a pop up saying the upload was successful.
- (7) You may start chatting with the subject.

4 User Study

The user study process starts with participants going through the Usage Instructions procedure above. We make sure the users get a sufficient product sense after chatting with a person they are familiar with. As they use it, we gather qualitative feedback based on their emotional reaction towards Ditto and remarks. After the initial demo period, we pose to participants these following study participant questions:

Study participant questions:

- (1) how much do you enjoy using large language models? (1-5)
- (2) are you comfortable using large language models? (1-5)
- (3) do you enjoy talking to or are more comfortable talking to ditto more than gpt? (y/n)
- (4) do you think you would learn concepts faster with ditto compared to gpt? (y/n)
- (5) do you see any future applications of this product?
- (6) what application would you use this for?
- (7) did ditto's language capabilities and usage mimic your person? (1-5)
- (8) do you see yourself using ditto to learn down the line? (y/n)

- (9) does ditto feel like the person it tried to imitate? (slang, general feel) (1-5)

- (10) how good are the language capabilities in general (1-5)

Broadly, these questions measure imitation ability, which is the realism of how well the agent imitates the person at hand. It also measures "usefulness", as discussed beforehand in the introduction. Moreover, importantly, after the user gets a good feel of the product, question 5-6 gauges the important use cases they see Ditto being applied towards.

We also add a fun activity for the users to do, which is a "turing test" where they try to trick the real subject by talking to them using Ditto's responses, and then seeing how long it takes for the subject to realize that it is not the real person talking. However, this is not a very reliable exercise as it requires for the other user to be active at the specific time of the user study, so this is just left as a thought experiment. In the future, we think we can design more effective scenarios to measure this, perhaps in-game.

5 Results

In general, Ditto was found to be both fun, engaging, and stimulating to be talking to someone familiar. Ditto is described to be talking "in the way the subject talks, but with GPT4's brain."

Question	Average Score (1-5)	Yes (%)	No (%)
Enjoyment of using large language models	4.6	-	-
Comfort with using large language models	4.5	-	-
Prefer Ditto over GPT	-	80	20
Learn concepts faster with Ditto	-	70	30
Ditto mimics the person	4.7	-	-
Use Ditto to learn in the future	-	75	25
Ditto feels like the person it imitates	4.8	-	-
General language capabilities of Ditto	4.7	-	-

Table 1: User study results for Ditto amongst n=15 participants

The study indicates a positive reception towards Ditto, with participants generally finding it easier and more enjoyable to talk to than GPT. On average, the participants rated their enjoyment of using large language models at 4.6 and their comfort level at 4.5. A significant majority (80%) preferred talking to Ditto over GPT, and 70% believed they could learn concepts faster with Ditto. Participants rated Ditto's ability to mimic the person at 4.7, indicating high realism in language imitation. Furthermore, 75% of participants expressed interest in using Ditto for learning in the future. The language capabilities of Ditto were rated highly, with an average score of 4.7.

Our application uses can be clustered into four use cases so far:

- (1) **Talking to somebody inaccessible.** This includes people like who have passed away and people you miss.

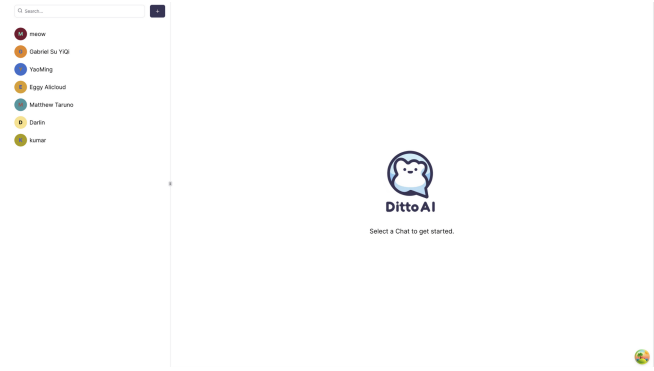


Figure 5: This is the Ditto homepage.

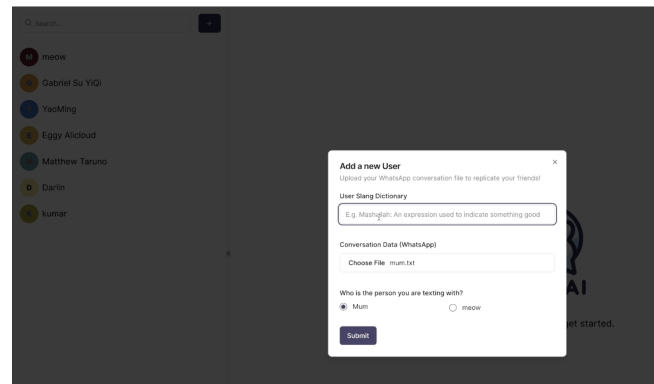


Figure 6: Upon clicking the + icon in the top-left, here is the upload flow pop-up for the user to upload the WhatsApp conversation data. The WhatsApp names are auto detected and the user can directly choose who they want to imitate, either subject 1 or subject 2. A user may also specify unique slang that they want Ditto to understand under the User Slang Dictionary field.

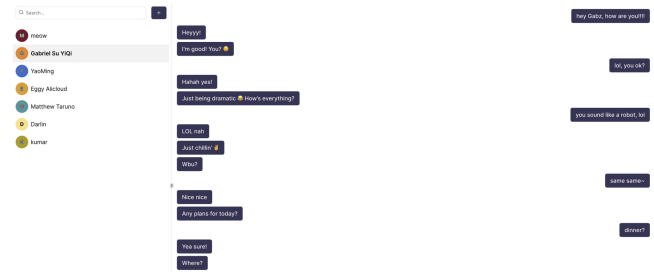


Figure 7: This is a sample trajectory conversation example with Gabriel.

- (2) **Introspection.** A tool used to understand ourselves better. In seeing and reflecting on how it imitates ourselves, we can come to learn more about ourselves.
- (3) **Educative learning assistant.** Learning from somebody you enjoy learning from, like a father figure.

- (4) **NPCs.** Replace digital world NPCs to talk more realistically. A test idea is to experiment Ditto on an in-game platform and apply a turing test in that environment.

6 Discussion

Here are the key insights from extensive usage and user testing of Ditto V1.

- (1) **Importance of Embedding Personality.** Upon analyzing many conversation trajectories, people are able to humanize Ditto more than for GPT4. This leads to expanded expectations, one major one being that participants could instantly recognize if the subject's personality deviated from the ground truth real person. In the current methodology we do not embed personality characteristics.
- (2) **Integrating Situational Awareness.** Another expectation people have over regular LM assistants is situational awareness. This could be things like the subjects' past relationships, food preferences, demographics, and opinions. To implement this we are able to use a memory architecture.
- (3) **Imitating self.** Since the conversation dataset typically includes yourself talking with another person, users tend to have the most of their own data. By means of high data availability, Ditto tends to be good at imitating the user. We tend to be very aware of our own tendencies, so when situational awareness and personality embedding issues arise, it was sensed very quickly.
- (4) **Multilingual abilities.** It is a reality for many people that everyday conversations are carried out with many loanwords and language switching within the same sentence. Adding the token distribution to the profile has helped Ditto exhibit appropriate multilingual usage in its conversations. One weakness we found amongst a trajectory example was that even though Ditto informed while cloning Darlin that she does not understand Chinese, upon talking Chinese to her, she still demonstrated that she knew what the Chinese meant in the answer.

7 Conclusion

In conclusion, Ditto has exceeded initial expectations in its ability to accurately mimic the messaging styles of subjects. The user study results indicate that Ditto not only provides a more relatable and engaging conversational experience compared to traditional language models like GPT-4 but also demonstrates strong potential across various applications. These applications include facilitating communication with inaccessible individuals, serving as an introspective tool for self-reflection, acting as an educational assistant, and enhancing the realism of digital NPCs.

We have shown that by incorporating profile elements such as emoji usage, slang, and conversational nuances as the core profile silhouette, Ditto creates a more human-like interaction that resonates with users on a personal level. There is a lot of room to expand these profile elements and even use more effective retrieval techniques and optimization for finding the right context to provide the foundation LM.

Future work will focus on expanding Ditto's capabilities by integrating additional conversational data sources (because more

conversation data leads to more accurate cloning) and refining the model's ability to capture and replicate diverse communication styles such as mid-sentence language switching and a deeper understanding of slang. There is also potential for developing more robust personality and situational awareness features to further enhance the model's realism and applicability.

The promising results from this study suggest that personal language models could play a crucial role in advancing the field of personalized language models, paving the way for more sophisticated and human-like AI assistants, and eventually a world where AI integrates deeper into our society as independent agents.

Feel free to explore Ditto and contribute to its development by visiting the following repository: <https://github.com/mtaruno/type-like-you>.

References

- [1] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova Das-Sarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. 2022. Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback. *arXiv preprint arXiv:2204.05862* (2022).
- [2] Lisa Bardach, Robert M. Klassen, and Nancy E. Perry. 2022. Teachers' psychological characteristics: Do they matter for teacher effectiveness, teachers' well-being, retention, and interpersonal relations? An integrative review. *Educational Psychology Review* 34, 1 (2022), 259–300. <https://doi.org/10.1007/s10648-021-09614-9>
- [3] Minghao Li, Yingxiu Zhao, Bowen Yu, Feifan Song, Hangyu Li, Haiyang Yu, Zhoujun Li, Fei Huang, and Yongbin Li. 2023. API-Bank: A Comprehensive Benchmark for Tool-Augmented LLMs. *arXiv preprint arXiv:2304.08244* (2023). <https://arxiv.org/abs/2304.08244> Version 2.
- [4] Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. *arXiv preprint arXiv:2304.03442* (2023). <https://arxiv.org/abs/2304.03442> Version 2.
- [5] Ji-Lun Peng, Sijia Cheng, Egil Diau, Yung-Yu Shih, Po-Heng Chen, Yen-Ting Lin, and Yun-Nung Chen. 2024. A Survey of Useful LLM Evaluation. *arXiv preprint arXiv:2406.00936* (2024). <https://arxiv.org/abs/2406.00936> Version 1.
- [6] Projjal. 2023. Human Conversation Training Data. <https://www.kaggle.com/datasets/projjal1/human-conversation-training-data/data>
- [7] Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, and Guoyin Wang. 2023. Instruction Tuning for Large Language Models: A Survey. *arXiv preprint arXiv:2308.10792* (2023). Last update: March 12, 2024.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009