

REVIEW



A Danish Chatbot for Kaj Munk: Experiences in Developing a Chatbot for the Study of Kaj Munk's *Ordet*

David Jakobsen^{1,*} , Simon K. Pacis¹ , Sara F. Jalk¹, and Peter Øhrstrøm¹ 

¹ Department of Communication and Psychology, Aalborg University, Denmark

Abstract: This study presents a pilot project developing a chatbot that is capable of conversing regarding Kaj Munk's drama *Ordet* (*The Word*) within a Danish cultural framework. Building on the argument for a dedicated Danish language foundation model, the project explores both linguistic and cultural challenges in generating accurate and idiomatic Danish responses. Three systems based on LLaMA2 (with and without Danish fine-tuning) and LLaMA3 were compared through factual and interpretive dialogs with the play. By fine-tuning a Danish translation of the OASST2 dataset and applying retrieval-augmented generation (RAG) to the digital Kaj Munk archive, this study evaluates model performance based on correctness, fluency, and cultural adequacy. Results show significant improvements with RAG and fine-tuning yet persistent traces of English language structure and bias. Beyond linguistic accuracy, the analysis highlights the need for models trained on culturally grounded corpora that reflect Danish literary traditions shaped by authors such as Grundtvig and Munk. The findings illustrate both the potential and limitations of large language models as tools for literary interpretation and for sustaining national cultural identity in AI-mediated dialogs.

Keywords: Kaj Munk, LLM, LLaMA2, LLaMA3, Danish

1. Introduction

A classic requirement for an AI system is its ability to respond to a given text in a way that appears intelligent. For many years, researchers sought to develop systems that are capable of such interaction by studying classical linguistics, particularly grammar, to create models that could be implemented to address this challenge. This issue was central to the work carried out in the 1980s at the research environment at Aalborg University, which later led to the academic field now known as “communication and digital media.” For an example, see the book *In Memoriam Inger Lytje* [1]. This research resulted in many intriguing computational linguistics projects that contributed to the understanding of the enigmatic nature of language comprehension. However, more and more researchers began to realize that building such rule-based systems to engage in a dialog with a human in a way that appears fully intelligent and human-like seemed an entirely unrealistic goal [2]. Instead, a different paradigm has emerged as more promising over the past few decades when it comes to developing systems that can simulate language understanding. Recently, impressive results have been achieved with chatbots that operate based on generative AI (GAI) using Meta's large language model (LLaMA), potentially enriched with a collection of selected texts applied via retrieval-augmented generation (RAG) technology on Kaj Munk's play *Ordet* and secondary literature on it. In the following, we will examine the potential of this technology from an empirical perspective. Our approach is a pilot project aimed at developing a chatbot that is capable of conversing with the Danish playwright and theologian Kaj Munk (1898–1944) [3].

The digital Kaj Munk archive contains texts written by Kaj Munk and texts discussing his thoughts, life, and works [4]. The archive is expanding, and as of May 14, 2024, it includes 5,862 unique scanned documents, 2,881 of which have been transcribed and are available online.¹ It is evident, however, that we are in the earliest stages of implementing LLaMA in digital archives for authors, particularly for non-English-speaking authors. This presents specific challenges regarding the reliability of the responses given by a chatbot. We address three kinds of challenges here. The first challenge concerns the problem pointed out by Enevoldsen et al. [5], regarding the lack of a Danish language foundation model. LLMs are often referred to as foundation models and have huge amounts of pretrained data adaptable to multiple tasks. Related to this is the second challenge concerning the underrepresentation of Danish cultural values and canonical authors, and the need for models that reflect Denmark's welfare-oriented (“well-being”) societal perspective, in contrast to American cultural assumptions. Finally, the third challenge concerns how to avoid hallucinations when dealing with an author who is not well known in many places outside Denmark. We will outline the technical and principal problems concerning the first and then report on the empirical results that we have achieved using an LLaMA trained to answer questions in Danish through RAG.

2. A Danish Chatbot for Kaj Munk

Several challenges related to LLMs can be categorized as hallucination problems defined as “generated content that is either

*Corresponding author: David Jakobsen, Department of Communication and Psychology, Aalborg University, Denmark. Email: davker@ikp.aau.dk

¹ See <http://arkiv.kajmunk.aau.dk>.

nonsensical or unfaithful to the provided source content.” Huang et al. [6] identified two main issues associated with hallucinations. First, the problem arises when an LLM uses vague or outright fabricated sources, and second, when the reliability and robustness of its responses are questionable. The hallucination issue is particularly relevant when training a chatbot for a Danish author. The core problem here is the absence of an LLM trained on Danish data, a concern recently addressed by Enevoldsen et al. [5].

Although models that show promising results for smaller languages like Danish have been developed [7, 8], these models often contain assumptions and cultural biases that are likely to influence translation and raise question of fairness [9, 10]. There is indeed a growing concern that LLMs contain demographic, geographic, and temporal biases in natural language applications [11, 12]. There are many vulnerabilities for bias in the process of making an LLM-powered chatbot. This could be viewed as a version of the hallucination problem: because context is crucial for an LLM, one must question whether a chatbot’s representation of a Danish language reality can truly be considered reliable. As examples, consider the cultural differences between the United States—where gun ownership, contrary to Denmark, is common—and Denmark’s distinct perspective on a welfare state [5]. In addition, more relevant to our study of Kaj Munk’s *Ordet* concerns religion. Although USA, like Denmark is a country grounded on Christian values, Denmark is a Lutheran country with a quite important split between the more liberal tradition following N.F.S Grundtvig and the more conservative Inner Mission [13]. In general, Denmark stands out as a more secularized country compared to USA, where religion in a more explicit way remains a prominent public force.

Our interest lies in testing the extent to which the lack of a Danish language model poses a challenge in the context of Kaj Munk. We aim to investigate to what degree this can be mitigated by incorporating additional Danish language material into an English language foundation model.

2.1. Fine-tuning to Danish

To make the best possible use of a modest grant for conducting a pilot study, we chose to focus on fine-tuning a model using RAG, capable of engaging in a dialog regarding one of Kaj Munk’s plays: the famous *Ordet* (“The Word”) [14]. We selected this play to facilitate a comparison between our model and GPT-4o. As our base model, we initially chose LLaMA2. However, during our work, Meta released LLaMA3, allowing us to compare three models: LLaMA2, with and without fine-tuning for Danish, and LLaMA3. LLaMA can be further trained and fine-tuned to perform better in a particular domain. To fine-tune LLaMA2 for Danish, we made a Danish translation of OASST2 [15]—a dataset in English that consists of an initial prompt as a root node, potentially followed by multiple nested responses. These responses can, in turn, have further replies, creating a dataset of numerous question–answer pairs that have been manually reviewed for accuracy. An example of a conversation from OASST2 is presented below:

User prompt: I am new to this AI thing. Can you tell me what kind of tasks you can do?

AI response: I can give you answers to many questions. You can ask me to: - Explain a wide array of topics in simple words - Give fun ideas to decorate your room - Write code in many different programming languages. These are just a small selection of things I can help you with, so feel free to ask!

This AI response was evaluated by three independent reviewers as being a correct and desirable reply to such a prompt. However, fine-

tuning an English model with English question–answer pairs is of little use for us, so we opted to use both Google’s MADLAD [16] and the OPUS model from the Technology Research Group at the University of Helsinki [17], both AI translation models, to translate OASST2 from English to Danish. We compared the translation results by calculating a BLEU score [18] for both translations and then compared them against each other. It is important to note that because we did not have any other Danish translations beyond the output from OPUS and MADLAD, we calculated a relative BLEU score between the two. A BLEU score is a metric used to evaluate machine-translated text by comparing it with a human made translation. We then chose the translation with the highest BLEU score, which turned out to be that of Google’s MADLAD. Subsequently, we used QLoRA to fine-tune LLaMA2 with the MADLAD-translated OASST2 dataset. The results clearly show that the addition of a Danish language dataset improved LLaMA2’s ability to provide responses in Danish. We posed two questions to the three models regarding Kaj Munk’s play *Ordet*:

“How long did it take Kaj Munk to write the play *Ordet*?”

“In Kaj Munk’s *Ordet*, how old was Johannes when he lost his mind?”

The following are the responses that we received:

1) LLaMA2 without fine-tuning to Danish

“Oh, det var en god spørgsmål! Kaj Munk tog og skrev skuespillet “*Ordet*” i årene 1938–1939. Det tog ham omkring et år at skrive dette monumentale skuespil, der er en af hans bedst kendte og highest i anseelse blandt teatersynsvillin.” (“*Oh that was a good question! Kaj Munk went and wrote the play *Ordet* in the years 1938–1939. It took him around a year to write this monumental play, it was one of his best known and highest regarded among theater...*”)

It is clear that there is an underlying English context in these responses. A phrase like “den var en god spørgsmål” (“*that was a good question*”) gets Danish grammar wrong. In Danish, “**den**” is used with common gender nouns, and “**det**” with neuter nouns. This is part of the grammatical gender system in Danish, where nouns can be either common gender or neuter. This is a common mistake committed by people who do not have Danish as their first language. The phrase “Kaj Munk tog og skrev *Ordet*” (“*Kaj Munk went and wrote *Ordet**”) would never be said in Danish. The phrase “*highest i anseelse*” (“*highest in esteem*”) mixes Danish and English and would never be said by a native Danish speaker. A phrase like “*monumentale skuespil*” (“*monumental play*”) is also noteworthy for two reasons. First, although it could technically be said in Danish, it is unlikely to be used in that context. More interestingly, it reflects a modern way of speaking that does not align with the way Danish was spoken in the first half of the 20th century. This might be what one wants, but together with the other mistakes, one is left wondering whether depths of appreciation for the time and circumstances of Kaj Munk’s work are lost in translation together with correct Danish grammar. We will delve into this in more detail later. However, reliance on the English context becomes even more apparent in the next question, where the answer is given entirely in English:

“Oh, fine! I’m happy to help you with that! Johannes, the protagonist of Kaj Munk’s play *Ordet*, was born in 1860, and he lost his faith in God at the age.”

It therefore seems crucial to fine-tune LLaMA2 with a Danish dataset, if only to eliminate the use of English and English expressions. The results show a clear improvement, although they still reveal the underlying English context.

2) LLaMA2 with fine-tuning to Danish

“Det var i 1920, da han havde 35 år.” (“That was in 1920 when he had 35 years.”)

The answer is brief and 1) entirely in Danish, but 2) the phrase “da han havde 35 år” (“when he was 35 years old”) reveals that the translation still has English as its underlying context. This is also evident from the second answer, which, unlike LLaMA2 without Danish fine-tuning, responds in Danish rather than directly in English:

“I Kaj Munks “*Ordet*,” blev Johannes født i 1870 og mistede sin forståelse i 1925, da han var 54 år gammel.” (“In Kaj Munk’s *Ordet*, Johannes was born in 1870 and lost his understanding in 1925, when he was 54 years old.”)

This response could almost be said by a native Danish speaker, except for the phrase “mistede sin forståelse” (“lost his understanding”), which might stem from the English “lost his senses.” However, there was a clear improvement from LLaMA2 to LLaMA3, as can be seen from the answers that it provided to the same questions:

3) LLaMA3

“Ifølge teksten i Munkiana 37 (s. 4) skrev Kaj Munk skuespillet *Ordet* på blot få dage i slutningen af 1925, efter en tragisk hændelse, som han havde set som præst i Vedersø.”

It is noteworthy that LLaMA3—unlike both versions of LLaMA2—is capable of referencing data from RAG. Even though the other two models were also configured to do so, they clearly did not. Furthermore, it is evident that the answer is provided in Danish in a way closer to how a native Danish speaker might give it. However, two exceptions must be noticed: the double use of “som” (“as” or “that”) in the final sentence, and the strangeness of saying (in Danish) that one has seen a tragic event “man ser en tragisk hændelse.” It sounds rather clumsy and ambiguous to say so in Danish.²

The second answer is also clearly better than those provided by the two versions of LLaMA2.

“I Kaj Munks *Ordet* er der ikke direkte beskrivelse af Johannes’ alder, da han mistede forstanden. Derimod kan vi se, at Johannes’ person er præget af en lang række traumatiske oplevelser, herunder hans mor og fars død, da han var ung, og senere hans egen præstevigt...”³

Note the correct use of the phrase “da han mistede forstanden” (“when he lost his mind”) and the precise use of the expression “Derimod kan vi se, at Johannes’ person er præget af...” (“On the other hand, we can see that Johannes’ character is marked by...”), which is appropriate when discussing a character in a play. However, there is still a small error in “lang række traumatiske oplevelser” (“long series of traumatic experiences”). Although “en lang række” is quite correct, an “e” is missing at the end of “traumatiske” to make it plural.

2.2. Evaluating the need for Danish

It is interesting to assess the degree of uncertainty, ambiguity, errors, or outright lack of proper Danish usage in the three models. Conversely, there is clear fine-tuning occurring as more Danish data are added to the dataset. One gets the impression that more training on more material directly translates into better results. This might suggest that the concern raised by Enevoldsen et al. [5] is premature. Is it, in the long

run, a problem that models trained on mostly English will “carry assumptions and cultural biases that may seamlessly transfer between languages and culture?” [5]. We might simply reach a point, eventually, where fine-tuning is so refined that it will be impossible to find errors in an LLaMA use of Danish. The slight improvement from LLaMA2 without Danish fine-tuning to one with it already makes this clear, but the progress toward LLaMA3 is significantly better. Here, the direct use of English in response to a Danish question disappears, and it can use language as one would expect when discussing a drama. Furthermore, it is evident that phrases like “mistede sin forståelse” (“lost his understanding”) are corrected to “mistede sin forstand” (“lost his mind”). It seems likely that we will soon reach models that are so proficient in Danish that only experts in Danish will be discern a supposedly influence of an English-speaking culture. However, viewed from this perspective, we are not capturing the core of the problem. The real issue is that the Danish language has been, and is, shaped by deeply ingrained cultural factors. These factors not only pertain to our view on guns and the welfare state. They also include the deep influence that authors Søren Kierkegaard (1813–1855), H. C. Andersen (1805–1875), Niels Frederik Severin Grundtvig (1783–1872), Georg Brandes (1842–1917), and Kaj Munk (1898–1944) have had on the language and, through that, Danish culture. The Danish language is molded around these authors’ books, songs, essays, and sermons. The genuine concern must be how we develop a language model trained on the right data that include the thoughts of these and other Danish writers or whether we end up with a language model that relies on machine-translated chats like OASST2 and other easily accessible sources. Buyl et al. [19] investigated whether LLMs reflect the political ideology of their creators and concluded that they indeed do. They argued that regulatory efforts aimed at producing ideologically neutral language models may be potentially harmful, precisely because it is doubtful whether such neutrality can be achieved. We concur with the conclusion of this study: it is essential to pursue transparency regarding the foundational ideological framework guiding the development of language models [19]. There is a need for not only a foundation model trained in Danish but also one that is trained on the works of important Danish authors who have contributed to shaping and cultivating the Danish language.

3. An Empirical Study of a Chatbot’s Potential

What are the prerequisites for engaging in a dialog with a chatbot based on LLaMA and RAG technology if the requirement is that the dialog should be regarding a specific topic, and within this topic, that the conversation should appear both rational and well informed?

In practice, there is a considerable difference in how well such chatbots perform as dialog partners. Fundamentally, the issue can be described as a “hallucination problem,” understood as situations in which a language model generates content that, to varying degrees, is not factually accurate, is imprecise, or is outright misleading [6]. We find the concept of “hallucination” useful for describing such instances but do not, on that basis, attribute consciousness to language models. That would only be the case if the model itself were uncertain as to whether what it conveys is indeed true. Even when, from our perspective, a language model appears to correct itself, we concur with Bender et al. [20] in referring to them merely as “stochastic parrots”:

² We are grateful to an anonymous reviewer for seeing the strangeness in this.

³ The model cut the sentence short here.

Contrary to how it may seem when we observe its output, an LM is a system for haphazardly stitching together sequences of linguistic forms that it has observed in its vast training data according to probabilistic information regarding how they combine but without any reference to meaning: a stochastic parrot [20].

Because human users are ultimately the ones interested in truth, we need to fine-tune these systems so that they provide factual answers, thereby allowing us to verify the validity and truthfulness of their content. In this regard, we aim to explore various ways of using RAG to avoid hallucinations. If different systems are to be compared, this can be conducted by evaluating the system's responses to a series of selected factual questions, for which there is a broad consensus on how the questions should be answered.

A good example of a topic suitable for dialog is Kaj Munk's famous play *Ordet* (1925), which is also well known from Dreyer's film adaptation. To assess selected GAI systems as partners in a dialog with a human regarding the play, we preselected a series of factual questions, which we asked three chosen systems to answer. The list of selected questions can be found in the Appendix. The 40 questions on the list are of a nature that would generally be expected to be answerable correctly by most readers familiar with the play *Ordet* and with some knowledge of its background. Provided that the list of questions was formulated independently of interaction with the systems, it is reasonable to use the percentage of correct answers as at least an initial measure of the systems' ability to serve as rational and well-informed partners in a dialog regarding *Ordet*.

To reinforce this independence, two of the authors (David and Peter) were responsible for the selection of the systems, and two others (Sara and Simon) were responsible for evaluating the systems' answers. The 40 answers given by each system were rated with 1 for correct, 0.5 for partly correct, and 0 for wrong. The investigations of the systems' responses were carried out in the spring and summer of 2024 using Aalborg University's supercomputer (HPC), which is available for research purposes. The systems can be constructed in various ways. Table 1 lists the three systems that we initially found interesting and therefore chose to examine more closely.

It is striking that System 3, which does not incorporate RAG technology, has a significantly lower score than the other two systems. In fact, none of the answers provided by System 3 was correct and only 15 were partly correct. It is generally well known that in many cases, a system's performance as a dialog partner can be significantly improved by incorporating RAG technology. The low score for System 3 also suggests that the underlying LLaMA training did not include many texts related to Kaj Munk's life and work. Clearly, it is crucial to include a selection of such texts if we are to build a system that is capable of engaging in a rational and well-informed dialog regarding the play *Ordet*.

Table 1
LLaMA3 systems

	Name of the System	Score
1	llama3-7b.6952,6953,6954,6955,933,6469_preprocess	0.7125
2	llama3-7b.933,6469,5775_rag	0.7375
3	llama3-7b.no_rag	0.1875

Note: The four digit numbers refer to text in the Kaj Munk archive of the journal *Munkiana*.

The difference between Systems 1 and 2 primarily concerns the selection of texts included in the preprocessing phase. Although there is only a slight difference between the overall results, it is worth noting

that System 2 gets 27 of the 40 questions correct and System 1 only gets 23 correct. It is worthwhile to look closer at the questions pertaining to the topic of miracles given its importance for understanding the play. Ten questions deal directly with miracles, namely, 23–25, 29, 30, and 33–37. System 1 has a score of 0.55, with four correct answers. System 2 scores markedly better, 0.85, with eight correct answers. In the following, we therefore choose System 2, which has the highest "factual" score among the mentioned systems, for a more detailed examination of the dialog capabilities of a system based on GAI. A high score on factual questions is certainly a positive indicator when aiming for a system that can serve as a well-informed dialog partner. Conversely, we have no guarantee that a high score on factual questions will also make the system appear intelligent when it comes to questions that do not rely on factual knowledge but rather on interpretation. Here, we can hardly speak of anything more than a working hypothesis—although most would likely find such a hypothesis quite reasonable.

In the next section, we will empirically investigate whether the selected system can also function effectively when the dialog shifts to themes that, for humans, require judgment, evaluation, and interpretation.

3.1 Various highlights from a debate with the chatbot

The chatbot reacts well when prompted to make a general characterization of the play:

User: Hvad handler Kaj Munks skuespil *Ordet* om? (*What is Kaj Munk's play Ordet about?*)

Chatbot: Kaj Munks skuespil *Ordet* handler om en konflikt mellem to livsanskuelser - et naturvidenskabeligt baseret livssyn og en understro. Det er et indlæg i livsanskuelsesdebatten og et opgør med datidens scientisme og ateisme. Stykket er centreret omkring skabelse, mirakel og under, og handler om en bonde, Mikkel Borgen, som på en gang styrer sit hus enevældigt som en anden patriarch og på den anden side ikke er bange for at erkende sin tvivl og usikkerhed på det religiøse område. (*Munkiana* 37, s. 14, <https://arkiv.kajmunk.aau.dk/documents/5775>) (*Kaj Munk's play Ordet is about a conflict between two worldviews—a scientific, naturalistic perspective and a superstitious belief. It is a contribution to the debate on worldviews and a critique of the scientism and atheism of the time. The play centers on themes of creation, miracles, and the supernatural, and revolves around a farmer, Mikkel Borgen, who, on one hand, rules his household with an almost patriarchal authority, yet, on the other hand, is not afraid to acknowledge his doubts and uncertainties in the realm of religion.*)

It seems that the system operates at a rather abstract level here and can incorporate new and thought-provoking perspectives. It should also be noted that the system was asked to provide references as documentation for the answers given. This is conducted quite neatly here. When the reference is checked, it also becomes clear where the ideas for the response originate—although it is not just a simple quotation.

One of the themes that we tried to get the chatbot to comment on concerns the worldviews or perspectives that clash in *Ordet*. This is particularly evident in the central scene of the play, where the elderly Borgen and Doctor Houen discuss their more principled approaches to life:

User: Hvad er forskellen på de livssyn som henholdsvis lægen Houen og gamle Mikkel Borgen forsvarer i skuespillet *Ordet*? (*What is the difference between the worldviews defended by Doctor Houen and the elderly Mikkel Borgen in the play Ordet?*)

Chatbot: Forskellen på de livssyn som henholdsvis lægen

Houen og gamle Mikkel Borgen forsvarer i skuespillet *Ordet* er, at Houen repræsenterer et rationelt og videnskabeligt syn på verden, mens Mikkel Borgen repræsenterer en grundtvigiansk tro, hvor Gud er menneskets velgører og tjener. Houen ser på verden med et rationelt og videnskabeligt syn, mens Mikkel Borgen ser på verden med en tro, der er baseret på følelser og intuition. (*The difference between the worldviews defended by Doctor Houen and the elderly Mikkel Borgen in the play Ordet is that Houen represents a rational and scientific perspective on the world whereas Mikkel Borgen represents a Grundtvigian faith, where God is seen as humanity's benefactor and servant. Houen views the world through a rational and scientific lens, whereas Mikkel Borgen's perspective is based on faith, emotions, and intuition.*) (Munkiana 37, s. 20, <https://arkiv.kajmunk.aau.dk/documents/5775>) (Munkiana 37, s. 7, <https://arkiv.kajmunk.aau.dk/documents/5775>)

Many would likely agree that the chatbot in the above response provides a fairly accurate characterization of the two worldviews. It should also be noted that the system again provides relevant references as documentation.

An equally well-informed response appears in the discussion regarding the character Johannes, who undergoes a significant transformation throughout the play:

User: Hvad sker der med Johannes i løbet af Skuespillet *Ordet*? (What happens to Johannes over the course of the play *Ordet*?)

Chatbot: I løbet af skuespillet *Ordet* sker der med Johannes, at han kalder sig Jesus og holder en bjergprædiken fra en klittop. Han repræsenterer den rene og stærke tro på Gud og er åbenbaringen af Guds vilje. Johannes' rolle i skuespillet er centralt, da han kalder sig Jesus og Jesu undergerninger er temaet i skuespillet. Han er den afgørende figur i det indre drama og er knyttet til Johannes Døberen og evangelisten Johannes. (*Over the course of the play Ordet, Johannes begins to call himself Jesus and delivers a sermon on the mount from a sand dune. He represents a pure and strong faith in God and is the manifestation of God's will. Johannes' role in the play is central, as he calls himself Jesus, and the miracles of Jesus are a key theme in the play. He is the pivotal figure in the internal drama and is associated with John the Baptist and the Evangelist John.*) (Munkiana 37, s. 7, <https://arkiv.kajmunk.aau.dk/documents/5775>) (Munkiana 37, s. 22, <https://arkiv.kajmunk.aau.dk/documents/5775>) (*Ordet*, s. 1, <https://arkiv.kajmunk.aau.dk/documents/933>)

The discussion regarding the possibility of miracles is central in *Ordet*. Here too, the chatbot engages in a reasonably sensible and well-informed manner:

User: Hvordan ser lægen Houen på den religiøse tro på miraklernes mulighed? (How does Doctor Houen view the religious belief in the possibility of miracles?)

Chatbot: Lægen Houen ser på den religiøse tro på miraklernes mulighed som en del af sin egen tro og videnskab. Han siger: "Selvfølgelig er jeg en Troens Mand; ellers var jeg en tarvelig Læge. Jeg har Tro til mig selv og min Videnskab og Tro paa de Mirakler, den har lært mig at gøre." Dette viser, at han ikke ser på troen på miraklernes mulighed som noget, der er i konflikt med hans videnskabelige forståelse, men snarere som en del af sin egen personlige tro og overbevisning. Han siger også: "Jeg gør Mirakler, og de bliver virkelig til noget," hvilket viser, at han ser på miraklerne som noget, der er muligt og reel. (*Doctor Houen views the religious belief in the possibility of miracles as part of his own faith and science. He says, "Of course, I am a man of faith; otherwise, I would be a poor doctor. I have faith in myself, in my science, and in the miracles it has taught me to perform." This shows that he does not see belief in the possibility of miracles as something in conflict with his scientific understanding, but rather as part of his own personal faith*

and conviction. He also says, "I perform miracles, and they truly come to pass," which indicates that he sees miracles as something possible and real.)

User: Er lægen Houen da religiøs? (Is Doctor Houen then religious?)

Chatbot: Lægen Houen i Kaj Munks drama "Ordet" repræsenterer et rationelt og videnskabeligt syn på verden. Han er en hovmodig og skråsikker mand, der fremstår som et kynisk, stoisk sandhedsvidne. Han er ikke religiøs i den forstand, at han ikke tror på mirakler og underet. Han ser på troen på Gud som noget, der er i konflikt med hans videnskabelige forståelse. Dette ses tydeligt, da han spørger Borgen: "Hvad tror De nu, der har hjulpet mest her i Aften: Deres Bøn eller min Præstation?" og da Borgen svarer: "Vorherres Velsignelse, kære Doktor, fra min Bøn til Deres Præstation, den har visst hjulpet mest." Lægen Houen er ikke religiøs i den forstand, at han ikke tror på Guds inddrejsning i verden.... (*Doctor Houen in Kaj Munk's drama Ordet represents a rational and scientific view of the world. He is an arrogant and self-assured man, appearing as a cynical, stoic witness to the truth. He is not religious in the sense that he does not believe in miracles and the supernatural. He sees faith in God as something that conflicts with his scientific understanding. This is evident when he asks Borgen, "What do you think has helped the most here tonight: your prayer or my efforts?" and Borgen responds, "The Lord's blessing, dear Doctor, from my prayer to your efforts, it certainly helped the most." Doctor Houen is not religious in the sense that he does not believe in God's intervention in the world.*)

The chatbot's contributions to the dialog on this central theme in the play's universe seem reasonably rational and well informed, although a bit unclear. Houen's views on religion and miracles could certainly have been expressed more clearly! Another surprising aspect of the above response is the expression "Guds inddrejsning i verden" ("God's entry into the world"), which is hard to interpret as anything other than a shortcoming in the chatbot's mastery of the Danish language.

4. Conclusion

The currently available systems that utilize GAI are quite fascinating. All indications suggest that with this technology, we may indeed succeed in building systems with which we can engage in a reasonable, well-informed, and even thought-provoking conversation regarding Kaj Munk's play *Ordet*. It appears that understanding the chosen literary text can—if not be automated—at least be simulated in a convincing way. There is no reason to doubt that something similar could be achieved with other literary texts. Conversely, it is also important to remember that the chatbot in question was specifically trained with texts regarding *Ordet*, and the requirement to train a chatbot on all Danish literary texts seems quite unrealistic. Of course, one can exhibit intelligence even if one is unaware of the specific topic of conversation, but it remains an open question whether there exists a form of general training for a chatbot that would allow it to pass a Turing test if the conversation could be regarding anything at all. A concern arises regarding the applicability of such systems for discussing an important Danish author like Kaj Munk. If the concern raised by Enevoldsen et al. [5] proves valid, a chatbot based on an LLM will reflect cultural contexts relevant to English rather than Danish. Our investigations seem to confirm the English foundation of LLaMA2, even when fine-tuned with additional Danish data. Although the Danish language performance improves in LLaMA3, the concern remains validated. However, the issue runs deeper, as confirmed by research on the ideological influence exerted by the creator of an LLM on the language model [19]. At the same time, it is essential to recognize that

literary interpretation is not static but historically contingent, subject to continual re-evaluation across time and cultural contexts. Any chatbot designed to engage with literature must therefore remain adaptable and responsive not only to new linguistic data but also to evolving interpretive frameworks. We propose that the way forward is to aim for not only a Danish language model but also one that intentionally and transparently includes the works of canonical Danish authors.

Ethical Statement

The authors declare that all potential respondents were fully informed about the survey and participation was voluntary.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in Figshare at <http://doi.org/10.6084/m9.figshare.30250096>.

Author Contribution Statement

David Jakobsen: Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Simon K. Pacis:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft. **Sara F. Jalk:** Validation, Formal analysis, Investigation, Data curation. **Peter Øhrstrøm:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft.

References

[1] Schärfe, H. (2003). *Impact: In memoriam Inger Lytje*. Denmark: Aalborg Universitetsforlag.

[2] Oppy, G., & Dowe, D. (2021). The Turing test. In E. N. Zalta & U. Nodelman (Eds.), *Stanford encyclopedia of philosophy*. Stanford University.

[3] Grund, C. (2023). An introduction to the theology of Kaj Munk. In S. Dosenrode & N. Mihaylova (Eds.), *Kaj Munk – An introduction to a controversial Danish priest, playwright, and debater in the inter-bellum* (pp. 35–72). Aalborg Universitetsforlag.

[4] Øhrstrøm, P. (2023). Interfaces to the digital Kaj Munk archive as research tools. In S. Dosenrode & N. Mihaylova (Eds.), *Kaj Munk – An introduction to a controversial Danish priest, playwright, and debater in the inter-bellum* (pp. 163–176). Aalborg Universitetsforlag.

[5] Enevoldsen, K., Hansen, L., Nielsen, D. S., Egebæk, R. A. F., Holm, S. V., Nielsen, M. C., ..., & Nielbo, K. (2023). Danish foundation models. *arXiv Preprint: 2311.07264*. <https://doi.org/10.48550/arXiv.2311.07264>

[6] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., ..., & Liu, T. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), 42. <https://doi.org/10.1145/3703155>

[7] Zhu, W., Lv, Y., Dong, Q., Yuan, F., Xu, J., Huang, S., ..., & Li, L. (2023). Extrapolating large language models to non-English by aligning languages. *arXiv Preprint: 2308.04948*. <https://doi.org/10.48550/arXiv.2308.04948>

[8] Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, 15(8), e12432. <https://doi.org/10.1111/llc3.12432>

[9] Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., & Hershcovich, D. (2023). Assessing cross-cultural alignment between ChatGPT and human societies: An empirical study. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP*, 53–67. <https://doi.org/10.18653/v1/2023.c3nlp-1.7>

[10] Ramesh, K., Sitaram, S., & Choudhury, M. (2023). Fairness in language models beyond English: Gaps and challenges. In *Findings of the Association for Computational Linguistics: EACL 2023*, 2106–2119. <https://doi.org/10.18653/v1/2023.findings-eacl.157>

[11] Qin, L., Chen, Q., Zhou, Y., Chen, Z., Li, Y., Liao, L., ..., & Yu, P. S. (2025). A survey of multilingual large language models. *Patterns*, 6(1), 101118. <https://doi.org/10.1016/j.patter.2024.101118>

[12] Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. *Journal of Data and Information Quality*, 15(2), 10. <https://doi.org/10.1145/3597307>

[13] Hall, J. A., Korsgaard, O., & Pedersen, O. K. (2015). *Building the nation: N. F. S. Grundtvig and Danish national identity*. Canada: McGill-Queen's University Press. <https://doi.org/10.1515/9780773596313>

[14] Munk, K., & Nielsen, N. (1932). *Ordet*. Denmark: Nyt nordisk forlag.

[15] Köpf, A., Kilcher, Y., von Rütte, D., Anagnostidis, S., Tam, Z. R., Stevens, K., ..., & Mattick, A. (2023). OpenAssistant conversations—Democratizing large language model alignment. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2064.

[16] Kudugunta, S., Caswell, I., Zhang, B., Garcia, X., Xin, D., Kusupati, A., ..., & Firat, O. (2023). MADLAD-400: A multilingual and document-level large audited dataset. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2940.

[17] Zhang, B., Williams, P., Titov, I., & Sennrich, R. (2020). Improving massively multilingual neural machine translation and zero-shot translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1628–1639. <https://doi.org/10.18653/v1/2020.acl-main.148>

[18] Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 311–318. <https://doi.org/10.3115/1073083.1073135>

[19] Buyl, M., Rogiers, A., Noels, S., Bied, G., Dominguez-Catena, I., Heiter, E., ..., & de Bie, T. (2024). Large language models reflect the ideology of their creators. *arXiv Preprint: 2410.18417*. <https://doi.org/10.48550/ARXIV.2410.18417>

[20] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>

How to Cite: Jakobsen, D., Pacis, S. K., Jalk, S. F., & Øhrstrøm, P. (2025). A Danish Chatbot for Kaj Munk: Experiences in Developing a Chatbot for the Study of Kaj Munk's *Ordet*. *Artificial Intelligence and Applications*. <https://doi.org/10.47852/bonviewAIA52025212>

Appendix:

40 Questions:

1 Hvornår skrev Kaj Munk skuespillet? *When did Kaj Munk write the play?*

2 Hvor længe var Kaj Munk om at skrive skuespillet? *How long did it take Kaj Munk to write the play?*

3 Hvorfor tilegnede Kaj Munk skuespillet til Hans Brix? *Why did Kaj Munk dedicate the play to Hans Brix?*

4 Hvem var Hans Brix? *Who was Hans Brix?*

5 Hvem bor i begyndelsen af skuespillet *Ordet* på Borgensgaard? *Who lives at Borgensgaard at the beginning of the play Ordet?*

6 Hvilken kirkelig retning bekender gamle Mikkel Borgen sig til? *Which church denomination does old Mikkel Borgen belong to?*

7 Hvad beder gamle Mikkel Borgen til Gud om? *What does old Mikkel Borgen pray to God for?*

8 Hvilken kirkelig retning bekender Peter Skrædder sig til? *Which church denomination does Peter the tailor belong to?*

9 Hvad er gamle Mikkel Borgen og Peter Skrædder uenige om? *What do old Mikkel Borgen and Peter the tailor disagree about?*

10 Hvem ejer Borgensgaard? *Who owns Borgensgaard?*

11 Hvem er gamle Mikkel Borgens ældste søn? *Who is old Mikkel Borgen's eldest son?*

12 Hvem er gårdens ældste søn, Mikkel, gift med? *Who is the eldest son of the farm, Mikkel, married to?*

13 Hvad hedder Mikkels og Ingers to døtre? *What are the names of Mikkel and Inger's two daughters?*

14 Hvorfor må Anders ikke få Anne, som han er forelsket i? *Why is Anders not allowed to have Anne, the one he loves?*

15 Hvad hedder Annes mor? *What is Anne's mother's name?*

16 Hvordan går det i skuespillet med Ingers graviditet? *What happens in the play with Inger's pregnancy?*

17 Hvordan går det i skuespillet med det barn, som Inger og Mikkel venter? *What happens in the play with the child that Inger and Mikkel are expecting?*

18 Mikkel Borgens ældste søn kommer til tro i skuespillet. Hvordan sker det? *Old Mikkel Borgen's eldest son comes to faith in the play. How does this happen?*

19 Hvor gammel var Johannes da han mistede forstanden? *How old was Johannes when he lost his mind?*

20 Hvordan viser det sig, at Johannes har mistet forstanden? *How does it show that Johannes has lost his mind?*

21 Hvad siger amtslæge Houen om mirakler? *What does district doctor Houen say about miracles?*

22 Hvad siger sognets præst, pastor Bandbul, om mirakler? *What does the parish priest, Pastor Bandbul, say about miracles?*

23 Hvilket stort mirakel sker sidst i skuespillet? *What great miracle happens at the end of the play?*

24 Hvad gør Johannes, da det går op for ham, at Inger er død? *What does Johannes do when he realizes that Inger is dead?*

25 Hvad sker der med Johannes mens han er væk fra gården og ude i klitterne? *What happens to Johannes while he is away from the farm and out in the dunes?*

26 Hvad gør Johannes, da han kommer hjem til gården, efter at have genvundet sin forstand? *What does Johannes do when he returns to the farm after regaining his sanity?*

27 Hvem opvækker sidst i stykket Inger fra de døde? *Who raises Inger from the dead at the end of the play?*

28 Hvem beder om, at Inger må opvækkes fra de døde? *Who prays that Inger may be raised from the dead?*

29 Hvordan reagerer Peter Skrædder på Ingers død? *How does Peter*

the tailor react to Inger's death?

30 Hvem leder Indre Mission i sognet? *Who leads the Inner Mission in the parish?*

31 Hvad mener pastor Bandbul om miraklernes tid? *What does Pastor Bandbul think about the age of miracles?*

32 Hvordan reagerer amtslæge Houen på Ingers genopvækelse? *How does district doctor Houen react to Inger's resurrection?*

33 Hvordan reagerer pastor Bandbul på Ingers genopvækelse? *How does Pastor Bandbul react to Inger's resurrection?*

34 Hvordan reagerer gamle Mikkel Borgen på Ingers genopvækelse? *How does old Mikkel Borgen react to Inger's resurrection?*

35 Hvordan reagerer Borgens søn, Mikkel, på Ingers genopvækelse? *How does Borgen's son, Mikkel, react to Inger's resurrection?*

36 Hvilke begivenheder i Munks liv kan have inspireret ham til at skrive skuespillet? *Which events in Munk's life may have inspired him to write the play?*

37 Hvorfor var skuespillets budskab vigtigt for Kaj Munk? *Why was the play's message important to Kaj Munk?*

38 På hvilket teater blev skuespillet først opført? *At which theater was the play first performed?*

39 Hvordan blev skuespillet modtaget af publikum? *How was the play received by the audience?*

40 Hvem filmatiserede skuespillet? *Who made the film adaptation of the play?*