Proposal details

Title of the project: Questions on social media as a window on our collective curiosity
Duration of the project (in months): 12

Public summary

English public summary (86/100):

Online media affect not only what we believe to be true or false, but also which questions we entertain, what we are curious or uncertain about. Such questions may, in turn, increase our vulnerability to disinformation: people more readily believe a lie if it answers a question they already had. This project will use the millions of questions posted on social media as a window onto our collective curiosity, to investigate, among other things, whether a surge in disinformation is preceded by changes in our curiosity.

Dutch public summary (80/100):

Online media beïnvloeden niet alleen wat we voor waar aannemen, maar ook welke vragen ons bezighouden, waar we nieuwsgierig naar zijn. Zulke vragen kunnen ons, op hun beurt, kwetsbaarder maken voor desinformatie: we trappen eerder in een leugen wanneer die een bestaande vraag lijkt te beantwoorden. Dit onderzoeksproject gebruikt de miljoenen vragen die we op sociale media posten als spiegel van onze collectieve nieuwsgierigheid, om onder meer te onderzoeken of een desinformatie-golf voorafgegaan wordt door een verandering in onze nieuwsgierigheid.

Budget

Type of costs	Clarification	Costs in euros
Research leave	0.6FTE, 6 months, PI	€ 23,194
Non-scientific personnel Acad	0.4FTE, 9 months (3+6), research assistants	€ 24,846
Travel and accommodation costs	Conference fee (4×)	€ 1,500
Total request from NWO		**€ 49,540**

Budget justification (105/150)

The PI will work on the project for the duration of one year, boosted in the first semester by a six-month research leave (0.6 FTE), and aided by student assistants for a total of 9 months (0.4 FTE), one for 3 months in the first semester, and another for 6 months throughout the entire second semester. Their respective tasks are outlined in the project proposal. We anticipate to present our results virtually (hence no additional travel cost) at three major conferences, at least one together with a student assistant (for reference, the ACL2022 conference virtual fee (inc. mandatory membership fee) was 375$ ≈ 360€).

Project proposal

Motivation. As citizens in the information age, our well-being and the health of our democracies depend more than ever on our own curiosity and on the questions we ask (Watson, 2022). Questions implicitly scaffold our conversations (Onea, 2016), they change how we process and remember incoming data (Grossnickle, 2016), and they can affect mental health and social cohesion (Kashdan et al., 2013). At the same time, our curiosity and doubts are constantly being manipulated by online media for short-term gain (e.g., click-bait) and long-term effects (propaganda, disinformation). Despite long research traditions on questions in the field of linguistics, on curiosity in education and cognitive science, and on disinformation in journalism and the social sciences, a lack of integration between these fields has resulted in crucial gaps in our scientific understanding. For instance, in linguistics, the frequent appeals to implicit questions are difficult to falsify in the absence of integration with curiosity research (Onea, 2016). In cognitive science, researchers often rely on questions as stimuli to trigger curiosity about a specific thing (Wagstaff et al., 2021), but these stimuli are designed on a case-by-case basis, without seeking a more general understanding of the relationship between questions and curiosity. In research on framing and rhetoric, sowing doubt has of course long been recognized as a propaganda tactic, but only in the form of mistrust in institutions and each other, not genuine uncertainty and curiosity about the world. Disinformation research, finally, primarily studies the impact of disinformation on our beliefs and opinions, not its impact on the questions we entertain, even though those questions could in turn render us more susceptible to disinformation that answers it, due to a cognitive bias known as *need for closure* (Webster & Kruglanski, 1994).

Objectives. Motivated by the latter possibility, i.e., that our questions might make us more susceptible to disinformation, as well as by the more general need for integration across disciplines, this project will develop and apply an innovative method to address two main research questions (RQs):

> **RQ1:** Prior to the spread and adoption of (dis)information on social media, is there an increase in the kinds of questions which this (dis)information answers? (Sub-question: Is there a difference between disinformation and non-disinformation in this regard?)

Although the answer to RQ1 is currently uncertain, one reason to expect an affirmative answer is that entertaining a question tends to make one more likely to accept the (dis)information that answers it, in line with the aforementioned bias *need for closure*. Another reason to expect an affirmative answer, for disinformation in particular, is that conspiracy theorists may adopt a facade of curiosity precisely to both exploit the aforementioned cognitive bias and to avoid accountability (*I'm only asking questions!*). The second research question is essentially a methodological generalization of the first:

> **RQ2:** Can we derive, from questions posted on social media, a meaningful, more general statistical-computational window on our collective curiosity and doubt?

We will try to answer this question by means of a constructive proof, building on the literature on describing semantic change (Xu & Kemp, 2015) and on natural language processing more generally in pursuing a *vector space* approach, detailed below. The resulting method will be tested and additionally motivated by applying it to a number of ancillary research questions, for instance to what degree trends in curiosity/doubt vary between sub-communities or change over time, whether changes can be correlated with world events (e.g., COVID vaccinations becoming available), and whether 'question bubbles' exist similar to (purported) information bubbles.

Method for RQ1 [months 1-4]: For both RQs we will use questions collected from social media posts as a new, valuable window onto our collective curiosity and doubt. Answering RQ1 additionally requires: 1) a way to select, with sufficient accuracy, social media posts containing disinformation and posts containing non-disinformation; and 2) a way to determine, with sufficient accuracy, whether a piece of (dis)information answers a given question. For the first component, we will follow the literature in combining blacklists of trusted and notorious sources with a filter based on hashtags (Giovanni et al., 2022). Note that we do not need to fully solve the difficult task of disinformation detection as we can simply ignore unclear cases: we merely want some disinformation, and some non-disinformation, not necessarily all (in decision-theoretic terms: precision matters, not recall). For the second component, we follow Qi et al. (2018) in adopting an off-the-shelf automated question answering system (fine-tuned if necessary on a sample of data we annotate) to determine whether a piece of (dis)information answers a given question. An affirmative answer to RQ1 would be supported if cross-correlational analysis (Zhang et al., 2018) revealed a time delay between two time series: the number of posts/shares/likes of a piece of information over time would tend to increase *after* an increase in posts/shares/likes over time of the questions it answers.

Method for RQ2 [months 5-12]: The main challenge for deriving a more general statistical window onto our collective curiosity and doubt, is the fact that functionally identical questions can be asked in very different ways (e.g., *What caused this? / Why? / How come?*), and, conversely, the same words can be used to ask different questions (*Who trusts this vaccine?* can refer to the COVID vaccine in one context, measles in another). What we need is a way to represent questions not by the literal sentence used to express it, but as much as possible by the underlying intent. For this we adopt the common method from the field of natural language processing of representing each piece of text – in our case questions from social media – as a high-dimensional *vector,* i.e., a long list of numbers. These numbers can be thought of as representing coordinates in a high-dimensional space (just like a point in two-dimensional space can be represented by two numbers: an x-coordinate and y-coordinate). This method lets us map questions onto vectors in such a way that similar questions end up close together in the space, henceforth **question space**. This common technique (e.g., Devlin et al., 2019) lets us abstract over superficial differences between questions and enables generalization based on (aspects of) the underlying intent, addressing the aforementioned challenge. To illustrate, one could imagine a question space constructed from COVID-related tweets where vaccine-related clarification questions are in one corner, causal explanation questions about the virus's origin in another, and questions about terminology like *reproduction number* and *mRNA* in yet another.

What exactly our question space looks like will depend on the precise way in which questions are mapped to vectors. T his project will compare and combine two common approaches for assigning vectors to pieces of text (in our case questions), namely a *feature-based approach* and *neural embedding-based approach*. The feature-based approach represents questions by a fixed number of predefined, theoretically motivated criteria, automatically extracted from the data with rule-based and/or machine learning-based algorithms, combining existing classification schemes (Graesser et al., 2008). The neural embedding-based approach does not use pre-defined features, but high-dimensional vector representations induced by artificial neural networks, a powerful and widespread method. By today's standards, we will use an off-the-shelf, pre-trained model in the transformers family (e.g., BERT; Devlin et al., 2019). These are deep artificial neural networks that have been trained on the task of predicting a missing words in a text (masked language modelling). Because this task requires not only grammatical and semantic knowledge but also aspects of pragmatic intent, such models 'learn' to assign vectors to text that contain such information. Feature-based approaches are generally representationally poorer (limited to the pre-selected features) but more directly interpretable, while neural embeddings are representationally richer but their dimensions lack clear interpretation; the former can be more useful for testing hypotheses about specific types of questions, while the more holistic view of neural embeddings is more conducive to identifying trends in curiosity/doubt in general.

Crucially, from the distribution of questions from a given dataset over question space, we can estimate a probability density function over the space (O'Brien et al., 2016), enabling us to estimate the probability of finding different types of questions (i.e., certain regions in the space) in the given dataset. This estimated probability density function is what this project will call the dataset's **question profile**. We can compute question profiles for instance for a social media platform as a whole (that is, a representative sample; e.g., Twitter's sample streaming API; Campan et al., 2018), for posts on a given topic or with a certain hashtag (e.g., *#covidiots*), or for all posts/likes/shares in some densely connected sub-network, i.e., a potential curiosity bubble. We can also compute question profiles at different moments in time, to see whether changes in curiosity correlate with world events. Since question profiles are probability densities, differences can be quantified by Kullback-Leibler divergence (i.e., relative entropy; for an application of this measure to similarly linguistically induced spaces see Xu & Kemp, 2015). This measures the degree to which the probabilities assigned to the various types of questions change when moving from one question profile to another. Comparing question profiles through network structure and through time will let us identify trends in curiosity/doubt and address the secondary research questions listed above (under RQ2), as well as further corroborating an affirmative answer to RQ2 itself.

Team and timeline. The project is divided into a shorter phase 1 devoted to RQ1 (4 months), and a longer phase 2 devoted to RQ2 (8 months). The PI will lead this project for the full one-year duration, with most advances made during the first six months, enabled by a research leave. Each phase of the project is aided by a research assistant (2 days a week for 3 months and 6 months, respectively). In phase 1 the assistant will iteratively develop and evaluate the (dis)information filters to be used, while the PI will collect and combine new and existing datasets (some already in our possession), adopt and apply a suitable question-answering model and conduct cross-correlational analysis of the resulting time series. In phase 2 the PI will develop the main modelling pipelines (months 5-6) of the feature-based and embedding-based approach, with the remainder of the project (months 7-12) being devoted to iterative evaluation and refinement as well as hypothesis testing, during which the division of labour between PI and assistant will be coordinated depending on the assistant's background and interests. This plan will enable us to establish the questions posed on social media as a valuable new window on our collective curiosity and doubt, a crucial resource – and liability – in the information age.

Literature references

- Campan, A., Atnafu, T., Truta, T. M., and Nolan, J. (2018). Is data collection through twitter streaming api useful for academic research?. In *2018 IEEE international conference on big data* (pp. 3638-3643). IEEE. https://doi.org/10.1109/BigData.2018.8621898
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, pages 4171–4186. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/N19-1423
- Giovanni, M. D., Pierri, F., Torres-Lugo, C., and Brambilla, M. (2022). Vaccineu: covid-19 vaccine conversations on twitter in French, German and Italian. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):1236–1244. https://doi.org/10.1609/icwsm.v16i1.19374
- Graesser, A., Rus, V., & Cai, Z. (2008). Question classification schemes. In *Proc. of the Workshop on Question Generation* (pp. 10-17). Retrieved November 2022 from https://www.cs.memphis.edu/~vrus/questiongeneration/16-GraesserEtAl-QG08.pdf
- Grossnickle, E. M. (2016). Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review*, 28(1):23–60. https://doi.org/10.1007/s10648-014-9294-y
- Kashdan, T. B., Sherman, R. A., Yarbro, J., and Funder, D. C. (2013). How are curious people viewed and how do they behave in social situations? From the perspectives of self, friends, parents, and unacquainted observers. *Journal of personality*, 81(2):142–154. https://doi.org/10.1111/j.1467-6494.2012.00796.x
- O'Brien, T. A., Kashinath, K., Cavanaugh, N. R., Collins, W. D., and O'Brien, J. P. (2016). A fast and objective multidimensional kernel density estimation method: fastKDE. *Computational Statistics & Data Analysis*, *101*, 148-160. https://doi.org/10.1016/j.csda.2016.02.014
- Onea, E. (2016). *Potential questions at the semantics-pragmatics interface.* Current Research in the Semantics/Pragmatics Interface, Volume: 33. Brill. https://doi.org/10.1163/9789004217935
- Qi, P., Zhang, Y., and Manning, C. D. (2020). Stay hungry, stay focused: Generating informative and specific questions in information-seeking conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 25–40, Online. Association for Computational Linguistics. http://dx.doi.org/10.18653/v1/2020.findings-emnlp.3
- Wagstaff, M. F., Flores, G. L., Ahmed, R., and Villanueva, S. (2021). Measures of curiosity: A literature review. *Human Resource Development Quarterly*, 32(3):363–389. https://doi.org/10.1002/hrdq.21417
- Watson, L. (2022). Cultivating curiosity in the information age. *Royal Institute of Philosophy Supplements*, 92:129–148. https://doi.org/10.1017/S1358246122000212
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.67.6.1049
- Xu, Y., & Kemp, C. (2015). A Computational Evaluation of Two Laws of Semantic Change. In *CogSci*. Retrieved November 2022 from https://www.cs.toronto.edu/~yangxu/xu_kemp_2015_parallelchange.pdf
- Zhang, Z., Zhang, Y., Shen, D., and Zhang, W. (2018). The cross-correlations between online sentiment proxies: Evidence from Google Trends and Twitter. *Physica A: Statistical Mechanics and Its Applications*, 508, 67-75. https://doi.org/10.1016/j.physa.2018.05.051