Designing and Evaluating an LLM-based Health AI Research Assistant for Hypertension Self-Management; Using Health Claims Metadata Criteria

LUUK P.A. SIMONS, PRADEEP K. MURUKANNAIAH,
MARK A. NEERINCX
Delft University of Technology, Faculty of EEMCS, Delft, Netherland
l.p.a.simons@tudelft.nl, P.K.Murukannaiah@tudelft.nl, M.A.Neerincx@tudelft.nl

Hypertension is a condition affecting most people over 45 years old. Health Self-Management offers many opportunities for prevention and cure. However, most scientific health literature is unknown by health professionals and/or patients. Per year about 200.000 new scientific papers on cardiovascular health appear, which is too much for a human to read. Hence, an LLM-based Health AI research assistant is developed for mining scientific literature on blood pressure and food. A user evaluation was conducted with n=8 participants who just completed an intensive lifestyle intervention for blood pressure self-management. They highlighted several challenges and opportunities for a Health AI, especially regarding claim transparency, data quality and risks of hallucinations. In the discussion we propose seven criteria using metadata and information characteristics to help evaluate ambiguous or conflicting health science claims.

Keywords:
hypertension, Self-Management support, decision support, eHealth, AI, LLM, claims analysis, metadata, public health

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.4.2024.16
ISBN 978-961-286-871-0

1 Introduction

When searching Google Scholar for 'cardiovascular health' more than 6.000.000 studies show up. In the years 2023 and 2022 the number of scientific publications on this topic referenced by Google Scholar are 249.000 and 307.000 respectively, per year. These numbers are staggering and impossible to keep up with for human intelligence. Hence, we are looking at AI (Artificial Intelligence) tools for assistance. Our *research goal* is to design a Health Research Assistant AI tool (or short "Health AI") based on LLMs (Large Language Models) with added metadata analysis tools.

The use case domain for this paper is 'hypertension and food.' One the one hand, hypertension affects roughly half the people above the age of 45 years in the developed world (Ostchega, 2020, Zhou, 2021, Carey & Whelton, 2020), but hypertension can be significantly reduced with healthy food choices (Roberts & Barnard, 2005, Franzini, 2012, Rodriguez-Leyva, 2013, Kapil, 2015, Dickinson, 2014, Siervo, 2015). On the other hand, we had the opportunity to conduct a Health AI user evaluation with a group of participants who had just completed the first two intensive weeks of a hypertension health behaviour self-management Challenge, giving them ample experience with hypertension information needs and various alternative sources of information, for us to investigate the added value of a Health AI concept.

Since this is a design paper, we first collect user feedback on a Health AI concept (section 4. Results), followed by a suggestion for seven metadata criteria that may be used for additional tools to help evaluate conflicting claims (section 5. Discussion).

Hence the *focus of the user evaluation* in this paper is: *'For users with hypertension health self-management experience, what are perceived usefulness and intention to use for a Health AI, compared to their other health information sources?'*

The Research Questions for the user evaluation are:

1. *In users' solution space, what are their information needs and priorities? What would they most want to ask the Health AI tool?*
2. *How do they use and value other information sources (besides the Health AI)?*
3. *What is their 'Technology Acceptance' evaluation and intention to use the Health AI?*

2 Related work

In this section we briefly address three topics: LLMs for health, claims analysis in AI (Guo, 2022), and the role of competing alternatives when designing and evaluating the added value of new tools.

In recent years multiple papers have been published on using LLMs for healthcare. Some using a review on opportunities and risks from mostly editorials (Sallam, 2023) or testing several use cases with health professionals (Cascella, 2023) or interviewing health professionals versus surveying the general public on their ChatGPT use in health (Raina, 2024). Some of the benefits that are relevant to our research question and generally mentioned in these papers are: utility in health research and benefits for health care practice (improving health literacy and efficiency in reviewing the literature). Risks that are often mentioned are: lack of transparency, risk of bias, incorrect citations, and risk of hallucinations. Other papers focusing more on the technology address privacy, security, and data architecture issues (Montagna, 2023) or training and evaluating specialised LLMs to increase natural language qualities like perceived helpfulness, logic and empathic phrasing (Lai, 2023). Overall, given that LLMs can be described as 'probable-word generators' (Shah, 2023), it is not so surprising that health care professionals describe their capabilities as lacking depth and argumentation in health expertise and lacking understanding of complex relationships between personal-, health- and behaviour-aspects (Raina, 2024).

However, we hold the view that from a technology perspective it is not enough to explicate risks of misinformation or lack of transparency of health claims. We must also think about the next steps forward: How to design and enhance generative AI tools such that these risks can be better managed? For example, when faced with conflicting claims from literature, it is not enough to just be transparent about the references used. To aid the user groups (see Method) we need to use metadata and develop additional tools that explicate how various sources and their claims can be weighed against each other. While using those tools, interpretation by human domain experts may likely be useful, hence creating a 'hybrid intelligence' (Simons, 2021, 2022a) combining the strengths of artificial and human intelligence. The specific domain of food and health has many conflicting claims (and conflicting interests of scientific authors). So, an important question is how to use metadata, information characteristics and assessment criteria to help evaluate claims.

The task of analysing claims is studied under the umbrella of *automated fact checking* (Guo, 2022) in the AI (specifically, Natural Language Processing) literature. Automated fact checking typically involves four subtasks: (1) Claim detection involves identifying claims for verification. An important aspect here is identifying claims that are check-worthy (i.e., claims whose truthfulness the public is interested in). (2) Evidence retrieval involves retrieving information which can be used to evaluate the veracity of the claim. (3) Verdict prediction involves determining the veracity of the claim by synthesising the pieces of evidence retrieved. (4) Justification production involves generating a justification for why a certain claim was ruled true or not true (or somewhere in between). This is an important and challenging task, considering the black-box nature of the AI tools. The main challenge for us is to formulate these tasks for the domain of our interest in a systematic manner.

Finally, we must borrow some value evaluation principles from the field of new product design. This paper reports on a user evaluation of a Health AI concept (see its description in section 3. Methods and Materials). Besides the general frameworks of TAM (Technology Acceptance Model, Venkatesh, 2000) and UTAUT (Unified Theory of Acceptance and Use of Technology, Venkatesh, 2003) looking at concepts like perceived usefulness and ease of use, product design aims to specify and design these qualities in detail (Rondini, 2016). Moreover, the added value of those new qualities should also be considered in comparison to competing alternatives (Herzwurm & Shockert, 2003, Rondini, 2016). Hence, in our user evaluation, besides asking feedback on perceived usefulness of various Health AI functions, we will also ask which other information sources are used and/or preferred. Previous research gave some indications for two alternative hypertension information sources (Simons, 2021). Firstly, when using Google Scholar to search on health interventions or disease causes, the results listed are overspecialized and too diverse. Plus, they are not very user-centred or action-oriented for health self-management. Secondly, when visiting sites of main health institutes, the latest science cannot be found and health advise is 'watered down' to not scare away the less health-conscious target groups (Simons, 2021, 2022a).

3 Methods and Materials

Research Design: In a design research approach (Vaishnavi & Kuechler, 2004, Verschuren & Hartog, 2005) we developed a high-level Health AI concept (see below) and collected feedback from n = 8 participants who had very recent experience with a healthy lifestyle intervention to reduce their hypertension, see section 4, Results. We used a mixed method approach for collecting user inputs and design suggestions: quantified surveys (see below), open questions and action research (during the intervention as well as the user evaluation) in the sense that we have a high level of 'access' to participants[1] while at the same time helping them navigate the information diversity they encounter. The user evaluations showed support needs for interpreting (sometimes ambiguous) claims. In section 5, Discussion, we propose several AI tooling options to support these user needs.

The Health AI concept:

- The initial scope is limited to the domain of food and blood pressure.
- Its training will include all (>100.000) recent scientific publications in this domain.
- It's precise details are yet to be determined, but its base is:
- For the user, the Health AI resembles ChatGPT, Bing and Bard, with the addition that it is specifically trained to help interpret recent studies.
- It works with questions and answers in plain Dutch. You can ask follow up questions on previous answers. You can see source publications which were used for the answers.

Participants: In the first week of February 2024, we collected feedback from n=8 Dutch participants which had started on January 15th with an intensive healthy lifestyle intervention to reduce their hypertension. All of them provided written consent. Details of this intervention are described elsewhere (Simons, 2022b, 2023a, 2024). Similar to previously published results, average blood pressures were successfully reduced from 140/87 to 122/77 in 12 days, thanks to many food- and other lifestyle improvements. The participants were all university employees: two of

[1] By supporting individuals during hypertension lifestyle interventions, as well as providing 6 months of healthy lifestyle coaching (Simons et al., 2010, 2017) for literally thousands of participants and caregivers in these domains, over the course of the past 10 years.

them scientists and six were supporting personnel. Half were male, half female. Their average age was 45, ranging from 29 to 58. All of them had experience with LLMs: two limited, three average and three of them a lot. All of them used multiple sources of information during the hypertension challenge.

User evaluation & data analysis: When going through health improvement iterations, participants cycle through three design spaces: 'problem'-, 'solution'- and 'evaluation space' (Simons, 2023b). This evaluation focusses on information usage for the 'solution space' (What are my most effective and attractive health behaviour options?) and the potential added value of the Health AI. In Table 1 the evaluation topics are listed. As explained in section 2, Related work, we need insights into their general information preferences (topic 1), which functionalities and support they would prefer from the Health AI (topic 2), their use of information sources during lifestyle changes, which are potentially competing options for the Health AI tool (topic 3), plus technology acceptance feedback (topic 4). For the latter, we used TAM (Venkatesh, 2000) and UTAUT (Venkatesh, 2003) for user evaluation concepts (perceived usefulness, ease-of-use, ability, trust, feeling, support, intention to use), with a focus more towards individual usage preferences, than on UTAUT's organisational technology adoption processes (Carlsson 2006). Regarding data collection and analysis, for each topic we used questionnaire items for quantified evaluation. We also asked users for additional inputs per topic on what they valued and why, given our design evaluation focus.

Table 1: User Evaluation topics

Topics information use and Health AI added value
1. Information usefulness, in general
2. 'Voice of the user' Health AI preferences
3. Use of other information sources (during self-management)
4. Technology acceptance' aspects for the Health AI

4 Results from the User Evaluation on the Health AI tool concept

To start our results section, we list the Research Questions (RQs) and the Tables containing the user evaluation summaries:

1. *In users' solution space, what are their information needs and priorities? What would they most want to ask the Health AI tool? (Table 2)*
2. *How do they use and value other information sources (besides the Health AI)? (Table 3)*
3. *What is their 'Technology Acceptance' evaluation and intention to use the Health AI? (Table 4)*

In Table 2, addressing the first RQ, we list user responses for information usefulness (7-point Likert scale). The first part of the table addresses their *general* information opinions, whereas the second part specifically addresses *Health AI* tool usefulness. We labelled the top 3 highest scores with green, for each question set, to highlight user preferences. Regarding general information usefulness, from the top 3 we can see that participants' priorities are learning which health behaviours help best for health/hypertension and how to make those changes easy. Question 1 was exceptional in that it gained maximum scores from everyone.

Other useful information sources mentioned (part 1 of Table 2, open question):

- The conversations with the coach were most useful. I would hope the AI could have a similar conversation with us.
- The context given during the Challenge in relation to healthy choices was very useful, like for example "how sugar- and saturated-fat-spikes heighten artery systemic inflammation".
- During the Challenge workshops we heard many things that you would never think of yourself, like for example the blood pressure lowering effect of seeds like flaxseeds.
- I was happy to hear about the updated hypertension guidelines from the AHA (American Heart Association), this is new for the Dutch context and I will include this in my conversations with my family physician.
- It's nice to see food intervention studies and effect sizes on hypertension.

Table 2: Information use & Health AI preferences (7-point (dis)agree, n=8, Avg=Average)

I find the following (general) information useful:	Avg Score
1. Connections between blood pressure, health and behaviour	7.0
2. Most effective behaviour changes for hypertension	6.4
3. Knowing blood pressure effect sizes of behaviour changes	6.0
4. Tips for making behaviour changes *easy*	6.6
5. Tips for making behaviour changes *successful*	6.1
The Health AI tool would be useful for:	
1. Comparing blood pressure effects of foods	5.9
2. Getting health feedback on a specific (supermarket) product	5.8
3. Learning the optimum dosage of a food product	5.0
4. Learning the broader health effects of a food	6.0
5. Comparing effect sizes of foods with other health behaviours	4.9
6. Practical tips on how to increase daily intake of health foods	5.8
7. Tips how to replace or avoid unhealthy foods	6.1
8. Tips how to deal with pitfalls/difficult moments	5.8

Part two of Table 2 shows the most useful applications of the Health AI tool, in the opinion of the participants. The top 3 scores are for learning blood pressure effects and broader health effects of food, plus practical tips on avoiding unhealthy foods. Just below the top 3 are three items each scoring 5.8 which all have a practical focus: daily eating patterns for increasing healthy foods, for dealing with pitfalls/difficult moments and aiding healthy choices when buying products in the supermarket. Interestingly, opinions varied on the practical advice items: some participants prefer to hear those practical tips from other participants (including usage/adoption context). As one of them stated:

"By interacting with the others about what works and why, our conversations really are part of our usage intention. The purpose is applying things yourself. ***Hence, the conversation is part of your own behaviour change****, instead of just information gathering."*

Still, others prefer the AI tool for practicality, versus preferring the coach for learning 'the bigger health picture and its relevant connections'.

Other Health AI usefulness mentioned (part 2 of Table 2, open question):

- It would be nice if the Health AI could filter information based on aspects like gender, age, weight, sports background, vegetarianism, etc, to increase relevance for my own situation.
- I would like to input my existing breakfast etc (which I like) and ask for health improvement suggestions.
- If certain foods are useful for my blood pressure, please show me the links to the original studies, so I can read them for myself. (See also Table 4)
- If the blood pressure food advice is distinct from the advice from my dietitian or weight watchers, can the Health AI explain why this may be so?
- I want to ask questions on other topics like aspirin or sauna: do they also influence my blood pressure?

In answer to RQ 2, Table 3 lists the extent of information source usage by participants during the Challenge period of two weeks. These can be viewed as alternative sources, potentially 'competing' with the Health AI we plan to introduce. Everybody indicated having used the coach inputs regularly and all participants except one said that the inputs from other participants were useful. The third most used source was formed by official health institutes. Regarding the fourth (= personal network), most indicated that this was more about bringing/sharing than about getting information; although information inputs were received from their network on the practicalities of implementing healthy lifestyle behaviours. Other Internet sources were explicitly labelled by most as containing too much confusing or low-quality information.

When asked what information was most useful (open question) all participants said that the ***Challenge workshops were most useful*** (including materials, PowerPoints, references, online portal with health information, plus the explanations provided). ***Reasons stated:*** *provided a good summary; value of the practical tips; a mirror to my own behaviours; the specific links and literature created a focused way for me to follow up on information; the summary and tips are saving me time; I don't feel the need to do my own research because this was good enough for me.*

Table 3: Use of other information sources (Number of times, n=8, Avg Nr=Average Number of times)

Number of times during Challenge (of 2 weeks)	Avg Nr
1. My personal network (family, friends, etc) [2]	1.7
2. My physician or other health professionals	0.4
3. Sites/info from official health institutes	2.3
4. Other Internet sources	0.5
5. Google Scholar, PubMed or similar	0.3
6. Individual scientific papers	0.9
7. Inputs/remarks from other Challenge participants	5.8
8. Inputs from Challenge coach	7.6

Regarding Research Question 3, Table 4 shows answers to various Technology Acceptance aspects. Since three of the highest scoring items have the same score (6.1) we labelled a top four of items green. From these it can be seen that on the one hand the Health AI is found interesting and there is an intention to use it. On the other hand it was clear (from items 2, 4 and 6, as well from the open answers) that all participants were wary about the risk of receiving unreliable answers from LLM tools like the Health AI. This is expressed in two of the top four items: 5. ('it will gain my trust, following the degree of clarity of its sources') and 7. ('I find it useful to discuss the Health AI outputs with the coach'). Being able to second-guess and interpret Health AI answers, especially using a human expert and hence creating a form of 'hybrid intelligence', is deemed a valuable way to use the Health AI. For this goal, interpretations of other participants (which have less expert knowledge in this domain) were deemed less useful.

On a practical level of ***anticipated future Health AI use, preferences varied*** (in line with the variation in Table 2 answers):

(1) some would prefer to get an introduction and practice session on how to (not) use it, whereas others would prefer to use it on their own,

[2] One of the participants was an outlier with score 15, hence excluded from this item average. Moreover, all participants said it was more about sharing information than receiving information, except for practical tips/discussions on how to implement health behaviours.

(2) some would like to be able to ask all the health, food and blood pressure questions they also asked in the workshops, others would focus on science mining, and still others would mainly want the Health AI for practical tips on daily health patterns (while bouncing these suggestions off others during workshop sessions).

Table 4: Technology acceptance factors (7-point (dis)agree, n=8, Avg=Average)

The Health AI tool..:	Avg Score
1. is interesting	6.1
2. is useful for insights on improving my health	5.5
3. is easy to use for asking questions	6.0
4. is easy to interpret when presenting conflicting articles	4.9
5. will gain my trust, following the degree of clarity of its sources	6.1
6. I find it useful to discuss its outputs with other Challenge participants	5.5
7. I find it useful to discuss its outputs with the Challenge coach	6.5
8. I find it useful to practice its use in Challenge workshops	5.8
9. I would certainly use the Health AI	6.1

5 Discussion & AI Tooling implications

A first limitation of this study is its explorative nature, with only n=8 participants. Still, for reaching input saturation at this design stage this appears sufficient; sometimes even five, six or seven users are enough (Faulkner, 2003). Second, the Health AI is only evaluated in concept. A next step in our research is to test a real prototype. Still, also on a concept level, user inputs are useful especially given their recent experience in dealing with ambiguous or conflicting claims from food and hypertension literature. Below we explore seven claim evaluation options, where metadata analysis tools can aid interpretation.

Criteria and information characteristics to evaluate ambiguous claims: Besides analysing claims structures in scientific literature, there are other metadata that can be used to help evaluate the reliability of claims. Table 5 lists several criteria against which (possibly contradicting) claims be evaluated.

Table 5: Claims evaluation criteria

Evaluation criteria & interpretation examples from literature:
1. *Time evolution of claims*: Tools that show claim changes over time can be useful. Dr Neal Barnard (2018) eloquently explains how claims on cardiac health of eggs have (incorrectly) become more positive in the past decades, exactly because the previous decades had been so complete and conclusive about the negative cardiac health effects. In short, 'serious research' moved elsewhere, leaving a void filled by the egg industry to create 'recent studies' with doubtful claims.
2. *Body of evidence*: As one of the most-cited scientists in the domain of health behaviours and health risks shows in an overview article (Willett, 2012), it is important to assess the extent of a body of evidence. For over a century, causal relationships between saturated fats, blood cholesterol and CVD (CardioVascular Disease) have been show by a broad array of (large scale) studies: from animal studies to prospective human migration studies across the world and large human RCTs (Randomized Controlled Trials).
3. *Consistency of claims*: An example of consistency is provided by all the studies showing health benefits of fruits and vegetables. Consistency in those findings is enormous. Still, some people (intervention participants, Internet sources, or sometimes even dietitians) discourage consuming more than two portions of fruits per day, claiming that their sugar content is bad for you. But even though refined sugars may be bad for you, when focusing on the overarching claims consistency (from studies) on fruits and their health effects, the consistency of positive health effects is clear and should prevail.
4. *Burden of proof*: Sometimes new claims go against 'prevailing wisdom.' This can either be a knowledge breakthrough (see next criterium) or a mistake. A famous example of the latter is the tobacco industry arguing that smoking is healthy since it reduces Parkinson's disease (Greger & Stone, 2016, p.265). Even though it is true that tobacco (and tomato) plants contain substances that infer Parkinson's protection, this is not enough. Burden of proof says that if there is massive proof pointing left (smoking kills you), then you need to carry a very heavy burden of proof for the opposite (smoking is healthy).
5. *Explicit arguments and proof for conflicting claims*: If you introduce a claim that goes against an existing *Body of evidence*, the *Burden of proof* is on you to give an explicit argument and/or proof why the new claim is valid, in the face of all other evidence. Soy health for humans is an example of this. In the past, we were faced with multiple animal studies showing cancer risks from high volume soy consumption (even though this was inconsistent with Asian populations consuming a lot of soy in good health). Finally, studies showed that the rodents in those animal studies metabolize soy differently from humans, and explained how the previous conflict in claims was resolved (Setchell, 2011).

Evaluation criteria & interpretation examples from literature:
6. Weighing claims for type of study: The previous soy example also illustrates an important fact in health: claims from a large scale, double-blind RCT in people carry much more evidence than animal studies (or observation studies). Even if this is obvious for some of us, it is helpful if a Health AI clarifies and uses this.
7. Claimer & industry affiliation analysis: Especially in the food sciences it is scary to see how many studies and scientists have industry affiliations and conflicting interests. For example, even in the US Dietary Guidelines Advisory Committee, where objectivity should be a priority, it turns out that 19 out of 20 members have clear industry affiliations and conflicting interests (Mialon, 2022). Hence, a metadata analysis on claimer identity & industry affiliations can provide useful insights to claim validity.

Hybrid intelligence for ambiguity 'Rationale capturing': In conclusion, our user evaluation confirmed the importance of information quality and science for healthy lifestyle choices. Especially regarding ambiguous or conflicting claims, participants expressed concern. They said they really valued support for interpreting those claims and all of them wanted to consult a human expert. Given the explanations we heard from these users, this finding appears to have external validity for other health topics as well.

In summary, using expert opinion to provide a 'rationale' behind confusing claims is deemed very valuable. This helps answer the user requirements: 'How to interpret claims?' And 'Is there an underlying story to explain the ambiguity?' Hence a 'hybrid intelligence' solution appears useful. In this paradigm the AI tools help reduce the information overload on experts, but the final advice is based on human (expert) explanation for the main user questions on claims confusion in the food and hypertension domain.

Acknowledgements

This research was (partly) funded by the https://www.hybrid-intelligence-centre.nl/ a 10-year programme funded the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, grant number 024.004.022 and by EU H2020 ICT48 project ``Humane AI Net'' under contract $\# $ 952026.

References

Barnard R. J. (2018). How the Egg Industry Skews Science. YouTube, accessed 26-2-2024: https://www.youtube.com/watch?v=FyG8wr0gWIA.

Carey, R. M., & Whelton, P. K. (2020). Evidence for the universal blood pressure goal of< 130/80 mm Hg is strong: controversies in hypertension-pro side of the argument. *Hypertension*, *76*(5), 1384-1390.

Carlsson, C., Carlsson, J., Hyvonen, K., Puhakainen, J., & Walden, P. (2006). Adoption of mobile devices/services—searching for answers with the UTAUT. In *System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on* (Vol. 6, pp. 132a-132a). IEEE.

Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of Medical Systems*, *47*(1), 33.

Dickinson, K. M., Clifton, P. M., Burrell, L. M., Barrett, P. H. R., & Keogh, J. B. (2014). Postprandial effects of a high salt meal on serum sodium, arterial stiffness, markers of nitric oxide production and markers of endothelial function. *Atherosclerosis*, *232*(1), 211-216.

Franzini, L., Ardigo, D., Valtuena, S., Pellegrini, N., Del Rio, D., Bianchi, M. A., ... & Zavaroni, I. (2012). Food selection based on high total antioxidant capacity improves endothelial function in a low cardiovascular risk population. *Nutrition, Metabolism and Cardiovascular Diseases*, *22*(1), 50-57.

Faulkner, L. (2003). Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, *35*, 379-383.

Greger, M., & Stone, G. (2016). *How not to die: discover the foods scientifically proven to prevent and reverse disease.* Pan Macmillan, New York City.

Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, *10*, 178-206.

Herzwurm, G., & Schockert, S. (2003). The leading edge in QFD for software and electronic business. *International Journal of Quality & Reliability Management*, *20*(1), 36-55.

Kapil, V., Khambata, R. S., Robertson, A., Caulfield, M. J., & Ahluwalia, A. (2015). Dietary nitrate provides sustained blood pressure lowering in hypertensive patients: a randomized, phase 2, double-blind, placebo-controlled study. Hypertension, 65(2), 320-327.

Lai, T., Shi, Y., Du, Z., Wu, J., Fu, K., Dou, Y., & Wang, Z. (2023). Psy-llm: Scaling up global mental health psychological services with ai-based large language models. *arXiv preprint arXiv:2307.11991*.

Mialon, M., Serodio, P., Crosbie, E., Teicholz, N., Naik, A., & Carriedo, A. (2022). Conflicts of interest for members of the US 2020 Dietary Guidelines Advisory Committee. *Public Health Nutrition*, 1-28.

Montagna, S., Ferretti, S., Klopfenstein, L. C., Florio, A., & Pengo, M. F. (2023). Data decentralisation of LLM-based chatbot systems in chronic disease self-management. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good* (pp. 205-212).

Ostchega, Y., Fryar, C. D., Nwankwo, T., & Nguyen, D. T. (2020). Hypertension prevalence among adults aged 18 and over: United States, 2017–2018, NCHS, National Health and Nutrition Examination Survey, 2017–2018: https://stacks.cdc.gov/view/cdc/87559.

Raina, A., Mishra, P., & Kumar, D. (2024). AI as a Medical Ally: Evaluating ChatGPT's Usage and Impact in Indian Healthcare. *arXiv preprint arXiv:2401.15605*.

Roberts, C. K., & Barnard, R. J. (2005). Effects of exercise and diet on chronic disease. *Journal of applied physiology*, *98*(1), 3-30.

Rodriguez-Leyva, D., Weighell, W., Edel, A. L., LaVallee, R., Dibrov, E., Pinneker, R., ... & Pierce, G. N. (2013). Potent antihypertensive action of dietary flaxseed in hypertensive patients. *Hypertension*, *62*(6), 1081-1089.

Rondini, A., Pezzotta, G., Pirola, F., Rossi, M., & Pina, P. (2016). How to design and evaluate early PSS concepts: the Product Service Concept Tree. *Procedia CIRP*, *50*, 366-371.

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare* MDPI, Vol. 11, No. 6, p. 887.

Setchell, K. D., Brown, N. M., Zhao, X., Lindley, S. L., Heubi, J. E., King, E. C., & Messina, M. J. (2011). Soy isoflavone phase II metabolism differs between rodents and humans: implications for the effect on breast cancer risk. *The American journal of clinical nutrition*, *94*(5), 1284-1294.

Shah, N. H., Entwistle, D., & Pfeffer, M. A. (2023). Creation and adoption of large language models in medicine. *Jama*, *330*(9), 866-869.

Siervo, M., Lara, J., Chowdhury, S., Ashor, A., Oggioni, C., & Mathers, J. C. (2015). Effects of the Dietary Approach to Stop Hypertension (DASH) diet on cardiovascular risk factors: a systematic review and meta-analysis. *British Journal of Nutrition*, *113*(1), 1-15.

Simons, LPA, Neerincx MA, Jonker CM (2021). Health Literature Hybrid AI for Health Improvement; A Design Analysis for Diabetes & Hypertension, pp. 184-197, *34th Bled eConference*. June 27-30, Bled, Slovenia, Proceedings retrieval from www.bledconference.org. ISBN-13: 978-961-286-385-9, DOI: https://doi.org/10.18690/978-961-286-385-9 .

Simons, LPA, Neerincx MA, Jonker CM (2022a). Is Google Making us Smart? Health Self-Management for High Performance Employees & Organisations, *International Journal of Networking and Virtual Organisations,* Vol 27, No 3, pp.200-216. DOI: 10.1504/IJNVO.2022.10053605

Simons, LPA, Gerritsen, B, Wielaard, B, Neerincx MA (2022b). Health Self-Management Support with Microlearning to Improve Hypertension, pp. 511-524, *35th Bled eConference*. June 26-29, Bled, Slovenia, Proceedings retrieval from www.bledconference.org. ISBN-13: 978-961-286-616-7, DOI: 10.18690/um/fov.4.2022

Simons, LPA, Gerritsen, B, Wielaard, B, Neerincx MA (2023a). Hypertension Self-Management Success in 2 weeks; 3 Pilot Studies, pp.19-34, *36th Bled eConference*. June 25-28, Bled, Slovenia, Proceedings retrieval www.bledconference.org. ISBN-13: 978-961-286-751-5, DOI: 10.18690/um.feri.6.2023

Simons, LPA, (2023b). Health 2050: Faster Cure via Bioinformatics & Quantified Self; A Design Analysis, *International Journal of Networking and Virtual Organisations*, Vol 28, No 1, pp.36-52. **DOI**:https://doi.org/10.1504/IJNVO.2023.130957

Simons, LPA, Gerritsen, B, Wielaard, B, Neerincx MA (2024). Employee Hypertension Self-Management Support with Microlearning and Social Learning. *International Journal of Networking and Virtual Organisations*, Accepted.

Vaishnavi, V and Kuechler, W. 2004. Design Research in Information Systems. Last updated August 16, 2009 from http://desrist.org/design-research-in-information-systems

Venkatesh, V. and Davis, F.D. (2000). "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," Management Science, 46, 186-204.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pp. 425-478.

Verschuren, P., & Hartog, R. (2005). Evaluation in design-oriented research. *Quality and Quantity*, *39*, 733-762.

Willett, W. C. (2012). Dietary fats and coronary heart disease. *J of internal medicine*, *272*(1), 13-24.

Zhou, B., Perel, P., Mensah, G. A., & Ezzati, M. (2021). Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, *18*(11), 785-802.