Intelligent Provision of Tailored, Easily Understood, and Trusted Health Information for Patient Empowerment

Marco Alfano¹,⁴, John Kellett²,⁶, Biagio Lenzitti³ and Markus Helfert¹,⁴
d
¹Innovation Value Institute, Maynooth University, Maynooth, Co. Kildare, Ireland
²Department of Emergency Medicine, Hospital of South West Jutland, Esbjerg, Denmark
³Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italy
⁴Lero, Science Foundation Ireland Research Centre for Software, Ireland

Keywords: Digital Health, Patient Empowerment, Intelligent Agent, Tailored Health Communication, Artificial Intelligence, Responsible AI, Big Data, Machine Learning.

Abstract: Although digital transformation in healthcare is accelerating, there is still a disconnect between current healthcare, focused on disease management, and a more holistic approach that looks at the health and wellbeing of the whole person. The latter approach aims at empowering patients and other health information seekers by improving their comprehension of their health so that they can manage it better. Currently, few stand-alone applications for patient empowerment exist and they seldom help users to understand health information. Thus, health information seekers often interact with the Web through generic search engines, which often produce results that are overwhelming, too generic, and of poor quality. This paper shows how the use of Artificial Intelligence (AI) in a responsible way may provide patients and others with health higher quality information that empowers them to improve their health and wellbeing. It presents an AI engine that extracts health content from the Web and provides the user with health information that is relevant, trustworthy, and easy to understand. The AI engine has been used to create an Intelligent Empowering Agent (IEA) that dialogues with users in simple language to provide customised information on symptoms and diseases, which helps them form their own evidence-based opinion on health matters that concern them.

1 INTRODUCTION

Although digital transformation in healthcare is accelerating, “the existing care pathways and care models rely on prescriptive approaches of health professionals who assess and direct care and treatment for patients, rather than creating care models designed and informed by patients to achieve personal goals and health outcomes” (Snowdon, 2020; Snowdon et al., 2014). The latter approach aims at empowering patients by improving their comprehension of their health so that they can manage it better (Alfano et al., 2019; European Health Parliament, 2017; WHO, 2016). Patient’s health literacy, information-seeking behaviour, sense of meaning, shared decision-making, and self-care/self-management are among the most important elements that characterise person/patient empowerment (Bodolica and Spraggon, 2019; Cerezo et al., 2016; Fumagalli et al., 2015). To be empowered a person/patient must:

1. have the necessary knowledge and self-awareness to understand health conditions and treatment options;
2. be able to make informed and conscious health choices (i.e., decide);
3. actively manage, with or without advice from medical professionals, their health and wellbeing (i.e., act).

Currently, few stand-alone applications for patient empowerment exist (Snowdon, 2020; Bodolica and
Spraggon, 2019) and patients and other users often go on the Web to search for health information (Alfano et al., 2019; Finney Rutten et al., 2019). However, Web information is mainly obtained through generic search engines and is often overwhelming, too generic, outdated and of poor quality (Finney Rutten et al., 2020; Alfano et al., 2019). Although Artificial Intelligence (AI) could play an important role in health empowerment (Iatraki et al., 2018; Kondyalkis et al., 2013), it often empowers machines rather than people (i.e., self-diagnosis apps tend to be substitute doctors and keep patients as passive recipients; Davenport and Kalakota, 2019; Fast and Horvitz, 2017; Jiang et al., 2017).

This paper presents the principles of the responsible use of AI for person/patient empowerment and describes an AI engine that can be used to extract and process health content from the Web. It provides the user with health information tailored to their specific needs, which is simple to understand and trustworthy. This AI engine powers an Intelligent Empowering Agent (IEA) that dialogues with users in simple language and provides customized information on symptoms and diseases, which helps them form their own evidence-based opinions on whatever health matters concern them.

2 RESPONSIBLE AI FOR PATIENT EMPOWERMENT

Responsible Artificial Intelligence is the “practice of designing, building, deploying, operationalizing and monitoring AI systems in a manner that empowers people and businesses, and fairly impacts customers and society – allowing companies to engender trust and scale AI with confidence” (Responsible Artificial Intelligence Institute, 2021). Although Responsible AI, Ethical AI, and Trustworthy AI are often used interchangeably (Responsible Artificial Intelligence Institute, 2021), in the context of our research to enhance patient empowerment, we prefer the term Responsible AI.

The basic principles of Responsible AI in health are analysed in the WHO guidance on “Ethics & Governance of Artificial Intelligence for Health” (2021). It shows that new technologies that use AI hold great promise to improve diagnosis, treatment, health research and drug development. AI technologies can also support government run public health functions, such as disease surveillance and outbreak control, provided ethics and human rights are at the heart of their design, deployment, and use.

The “Ethics guidelines for trustworthy AI” by the EU High-Level Expert Group on Artificial Intelligence (HLEG, 2019), sets out a list of seven key requirements on the AI systems:

- Human agency and oversight;
- Technical robustness and safety;
- Privacy and data governance;
- Transparency;
- Diversity, non-discrimination, and fairness;
- Societal and environmental wellbeing;
- Accountability Trustworthy.

Responsible AI requires a holistic and systemic approach, encompassing the trustworthiness of all actors and processes that are part of the technical and socio-technical context. The most important actors are human beings, and they must be able to interact with AI systems at any stage. In particular, the first requirement, “Human agency and oversight”, states that AI systems should empower human beings by allowing them to make informed decisions and safeguarding their fundamental rights. Users should be given the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree and, where possible, be enabled to reasonably self-assess or challenge them. AI systems should support individuals in making better, more informed choices, in accordance with their goals (HLEG, 2019).

There cannot be truly Responsible AI in health without the direct intervention of “empowered” human beings (health professionals and patients) who must be able to understand both the AI-system process and its outcome, and directly participate in the decisions and actions that originate. For person/patient empowerment, the first step is understanding health information. A previously published research and a literature review (Alfano et al., 2019, 2021) identified the following metrics of health-information comprehension for empowerment purposes:

- language complexity;
- information quality
- information classification/customization (tailoring).

Thus, person empowerment becomes both a requisite and an outcome of Responsible AI (Fig. 1) and to have Responsible AI in health:

- AI systems must behave responsibly by applying ethical principles such as the ones discussed above (technical requirements);
- AI systems must help people to become empowered by providing health information
that presents characteristics that facilitate its comprehension (socio-technical requirements).

The application of these principles leads to the “virtuous circle” shown in Fig. 1. A Responsible AI system facilitates person empowerment and an empowered person exerts “human agency and oversight” on the AI system.

![Figure 1: “Virtuous” interaction between Responsible AI and an empowered person.](image)

### 3 USING AI FOR EXTRACTING AND CLASSIFYING WEB HEALTH INFORMATION

As shown in Alfano et al. (2020) and in Fox et al. (2013), the most searched type of health information on the Web is about medical problems and diseases. Therefore, we decided to provide the user with web information on symptoms and diseases organised in sections. This is similar to what many other health web sites do, but in a more structured, comprehensive, and interactive way. Initially, we made a visual analysis of fifty health web pages of symptoms and diseases to understand what kind of information is provided and how it could be best divided into sections. The headings of each web page, their semantic equivalents and themes were analysed to determine those which were most frequently provided; these were then grouped together into the following initial classification:

- **Overview** (summary, what is it, what are the types, definition, description, consideration, types, types of, deeper types of, basics, basic information).
- **Alternative name** (synonyms, substitutes).
- **Symptoms** (what are the symptoms, clinical presentation, what you feel, related complaints, what you cannot do anymore).
- **Causes** (what is main cause, it is hereditary, causes, possible causes, most common causes, most serious causes, what causes, other causes, culprit, causes of, health conditions).
- **Diagnosis and test** (how is it evaluated and diagnosed, what test will be done to diagnose it, diagnosing, what to expect at a medical visit, location of your pain, type and intensity, history of your pain, other medical history, other types of diagnosis, how is it diagnosed).
- **Risk factor** (who gets it, what are the risk factors, who’s at risk, who does it affect).
- **When to see a doctor** (when to contact professionals, when to contact a medical professional, when should I call my healthcare provider, when to see a doctor, what symptoms require medical care, when to worry, symptoms that require urgent care, when should I see my healthcare provider).
- **Management and treatment** (how is it treated, medication for it, treating the underlying condition, how can I get rid of it, home care, care and treatment, medication, medications, how to ease it yourself, treatment for, medical treatment, home treatment, surgery).
- **Prevention** (how can I prevent it, beware of, can it be prevented).
- **Outlook and prognosis** (can it be cured, what happen after I start treatment, what should I expect in the long term, complications, what is the outlook if you have it, what’s the outlook, living with).

Next, we created a list of 474 symptoms derived from a classic textbook on signs and symptoms (MacBryde, R.S. Blacklow, 1970), review of the literature, and expert opinion, and a list of 801 diseases derived by the commonest diagnoses encountered in primary care (Finley et al., 2018) and the diseases most often associated with in-hospital death (Kellett and Deane, 2007).

Machine learning was then applied to automatically classify the headings of web pages related to the symptoms and diseases to the classes seen above. This was done by using the following functions:

- **Url_Generator (k, n)** uses a Google™ API to retrieve n URLs of keyword k (e.g., a symptom). For each symptom and disease, 40 URLs were retrieved.
BaseScraper (URL) takes each URL and extracts all <h> headings.

MachineLearner uses a TensorFlow algorithm to create an AI model that allocates headings to classes. TensorFlow is an open source library for numerical computation and large-scale machine learning. It allows to create dataflow graphs—structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor (Abadi et al., 2016). The AI model has been trained with 1500 headings manually allocated to different classes, reaching a ~ 85% precision.

Sorter uses the AI model to classify headings by indicating the probability a heading belongs to a class. It uses the TextVectorization() function, provided by the Keras framework, to transform text into a vector by also removing capital letters and punctuation to avoid similar words counted as different vectors.

As a last step, for each symptom or disease, the 40 various headings, were extracted, using the DOM of downloaded. The HTML contents, related to the web pages indicated by Url_Generator were downloaded. The HTML contents, related to the various headings, were extracted, using the DOM of the HTML page, and stored in a DB. The most fitting information, for each symptom/disease section, is chosen based on:

### Information Relevance

The proximity of the symptom/disease with the title of the page, \( S_c \), is computed as the maximum of the cosine of similarity between the symptom/disease and its synonyms (A) and the title of the page (T), as follows:

\[
S_c(A,T) = \max \frac{\sum_{i=1}^{n} A_i T_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} T_i^2}}
\]  
(1)

where \( A_i \) and \( T_i \) are the occurrences of the same word present in the two texts and \( n \) is the number of words present in \( A \) and \( T \).

The symptom/disease occurrence index, \( I_{occ} \), is computed as the maximum of the ratio between the average occurrences of the symptom/disease and its synonyms (A) and the number of occurrences of the most frequent word in the page (P), as follows:

\[
I_{occ}(A,P) = \max \frac{n_i}{\sum_{i=1}^{n} A_i}
\]  
(2)

For improved results, we have applied a stemming process to all words and we have evaluated the similarities between words by using the normalized Levenshtein distance (Yujian and Bo, 2007).

### Information Quality

The quality of a web page \( P \) is computed by counting the meta data of the page that can be associated to the information quality (Alfano et al., 2021; WHO, 2003; Eysenbach, 2002) and grouping them in four categories:

- **Temporal indicators** \( T(P) \) (Last-Modified, revised, PublishDate, ...);
- **Spatial indicators** \( S(P) \) (og:email, og:phone_number, og:street-address, ...);
- **Descriptive indicators** \( D(P) \) (abstract, summary, description, og:description, ...);
- **Author indicators** \( A(P) \) (author, citation_author, ...).

A fifth indicator \( O(P) \) refers to the most popular health web sites.

The Information Quality, \( I_q \), is then computed as follows:

\[
I_q(P) = \frac{T(P)S(P)D(P)A(P)O(P)}{5}
\]  
(3)

### Language Complexity

The Language Complexity, \( L_c \), of a web page \( P \), is computed with the following formula:

\[
L_c(P) = \frac{\sum_{i=1}^{n} WF(W_i) A_i}{\sum_{i=1}^{n} A_i}
\]  
(4)

where \( W_i \) are the words of a text \( P \), \( WF(W_i) \) is the Word Familiarity, i.e., the number of Google results of \( W_i \) (Alfano et al., 2021; Kloehn et al., 2018, Leroy et al., 2016), and \( A_i \) are the occurrences of \( W_i \).

The most fitting information is computed by using the following weighted formula:

\[
FitInfo = \alpha \cdot S_c + \beta \cdot I_{occ} + \gamma \cdot I_q + \delta \cdot L_c
\]  
(5)

Where \( \alpha + \beta + \gamma + \delta = 1 \) and allow to differently weigh the information relevance, information quality

---

1 https://www.tensorflow.org/
and language complexity. A Support Vector Machine (SVM, Cortes and Vapnik, 1995) has been trained with human classified web pages to provide the best combination of the weights and then the most appropriate information for each section.

4 INTELLIGENT EMPOWERING AGENT

The AI engine, presented in the previous section, has been implemented as part of an Intelligent Empowering Agent (IEA) that provides health information tailored to the users’ needs, which is intelligible, current, accurate, trustworthy, valuable, and usable. The IEA model is shown in Fig. 2. Its components have been described in detail (Alfano et al., 2022) and can be summarized as follows:

a. **User Query.** The user selects a complaint from a list or directly enters it as free text.

b. **User Profile.** The user provides some information about him/her (currently age range and sex).

c. **Search Engine.** The search engine retrieves the top 40 Google results for each symptom/disease.

d. **AI Algorithm.** The AI algorithm uses the AI model shown above to select the health information according to following criteria:
   - **Custom Information**, to provide users with tailored content for the symptom/disease of interest, organized in sections.
   - **Information Quality**, to provide users with current, accurate, trustworthy, and unambiguous information.
   - **Language Complexity**, to provide users with information that they can easily understand.

e. **Output Presentation.** Tailored relevant health information is provided on complaints, diseases, medical tests, when to see a doctor, treatment, prevention, and scientific articles according to the section headings presented above. The scientific-articles section uses a PubMed™ API to retrieve the most relevant articles on the topic of interest.

   A “traffic-light” colouring coding (i.e., red, amber, or green), that implies the need for an urgent consultation with a healthcare professional, is also provided.

   In an IEA prototype, the Conversational Health Agent for Person Empowerment (CHAPE - http://cohealth.ivi.ie/chape/), users input their age and sex; CHAPE then provides a list of possible complaints that can be easily understood, such as pain or discomfort, breathlessness, and weakness or fatigue. Depending on the complaint selected and the user’s profile characteristics, a further sub-list of possibly related complaints is presented, to help define the primary complaint more precisely (Fig. 3).

   Alternatively, the user can directly type in any complaint in a free text area.

---

Figure 2: IEA model.

---

Figure 3: CHAPE interface allows users to specify their complaints in an easy and natural way.

An information window is then displayed (Fig. 4) and it contains:

- Complaint name with a background colour (red, amber, or green), which indicates the health risk.
- Complaint description.
- Alternate names of the complaint.
- Related complaints.
- Disease(s) associated with the complaint.
- Tests commonly used to further define the complaint.
- When to see a doctor
- Treatment
- Prevention
- Scientific articles

Figure 4: Output window containing information about the searched element.

When a related complaint, disease, or test is clicked on, a new information window for that element is opened. The list of related complaints, diseases and tests of the new window is ordered so that the elements that are in common with the previous searches are shown first (Fig. 5). This further customizes the provided information by following the user search path.

Some subjective tests have been carried out to assess the effectiveness of the IEA in terms of usability, user experience and perceived value (Alfano et al., 2022). In terms of usability, most respondents found CHAPE interface clear and helpful in identifying health information about complaints, diseases, and tests. In terms of user experience, most respondents could better understand their complaints and related diseases and tests. In terms of perceived value, the majority of respondents found that CHAPE would improve communication with their doctors. Expanding the user profile and improving the system’s interaction interface were the main recommendations.

5 CONCLUSIONS

This paper shows how the use of AI in a responsible way may help empower a person/patient better manage their health and wellbeing. It presents an AI engine that extracts health content from the Web to provide the user with health information that is relevant, trustworthy, and easy to understand. The AI engine powers a prototype Intelligent Empowering Agent (IEA) that dialogues with user in simple language and provides tailored, trustworthy information on symptoms and diseases, which help users to form evidence-based opinions on health matters that concern them. To our best knowledge, this is the first attempt to create an intelligent empowering agent that exploits the potential of AI and the vast amount of health information available on the Web to facilitate comprehension and action on general complaints/diseases.
6 FUTURE RESEARCH/DIRECTIONS

Although third party subjective assessments are encouraging, the user profile and the system interaction need to be improved. To this end, the user interface is being provided with a graphical representation of the body, for complaint identification and location, and more user profile information, such as gait, body type, nutritional status, comorbidities, are being added. Complaints and diseases are being associated with Concept Unique Identifiers (CUI) of the Unified Medical Language System (UMLS) to map them to standard terms taken from medical-term classifications such as ICD-9, ICD-10, or SNOMED. How information is gathered and filtered out will be improved and explicitly explained to improve trustworthiness. Although user input is anonymous, users will be provided with an option to grant or withdraw informed consent to use their data. Finally, the agent is going to be tested on a wider demographic.

ACKNOWLEDGEMENTS

This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094 P2 and co-funded under the European Regional Development Fund through the Southern & Eastern Regional Operational Programme to Lero - the Science Foundation Ireland Research Centre for Software (www.lero.ie). Thanks to Yvan Pannefelieu, from ESEO Grand École d’Ingénieurs, for his contribution on the machine learning algorithm for the extraction and classification of HTML headings.

REFERENCES


10 https://www.cdc.gov/nchs/icd/icd9cm.htm
11 https://www.cdc.gov/nchs/icd/icd-10-cm.htm
12 https://www.snomed.org/
Intelligent Provision of Tailored, Easily Understood, and Trusted Health Information for Patient Empowerment


