Hybrid Collective Intelligence for Decision Support in Complex Open-Ended Domains

Vito TRIANNI [a,1], Andrea Giovanni NUZZOLESE [a], Jaron PORCIELLO [ag],
Ralf H.J.M. KURVERS [b], Stefan M. HERZOG [b], Gioele BARABUCCI [cf],
Aleksandra BERDITCHEVSKAIA [d], and Fai FUNG [e]

[a] *ISTC-CNR, Italy*
[b] *Max Planck Institute for Human Development, Germany*
[c] *Human Dx EU, Ireland*
[d] *NESTA, UK*
[e] *MET Office, UK*
[f] *Norwegian University of Science and Technology, Norway*
[g] *University of Notre Dame, IN, USA*
ORCiD ID: Vito Trianni https://orcid.org/0000-0002-9114-8486, Andrea Giovanni
Nuzzolese https://orcid.org/0000-0003-2928-9496, Jaron Porciello
https://orcid.org/0000-0002-3179-1971, H.J.M. Ralf Kurvers
https://orcid.org/0000-0002-3460-0392, M. Stefan Herzog
https://orcid.org/0000-0003-2329-6433, Gioele Barabucci
https://orcid.org/0000-0003-4586-0111, Aleksandra Berditchevskaia
https://orcid.org/0000-0003-0325-9950, Fai Fung
https://orcid.org/0000-0003-4367-151X

Abstract. Human knowledge is growing exponentially, providing huge and sometimes contrasting evidence to support decision making in the realm of complex problems. To fight knowledge fragmentation, collective intelligence leverages groups of experts (possibly from diverse domains) that jointly provide solutions. However, to promote beneficial outcomes and avoid herding, it is necessary to (i) elicit diverse responses and (ii) suitably aggregate them in a collective solution. To this end, AI can help with dealing with large knowledge bases, as well as with reasoning on expert-provided knowledge to support decision-making. A hybrid human-artificial collective intelligence can leverage the complementarity of expert knowledge and machine processing to deal with complex problems. We discuss how such a hybrid human-artificial collective intelligence can be deployed to support decision processes, and we present case studies in two different domains: general medical diagnostics and climate change adaptation management.

Keywords. Collective intelligence, decision support systems, knowledge graphs

---

[1] Corresponding Author: Vito Trianni, vito.trianni@istc.cnr.it

1. Introduction

Humanity is facing complex problems that require engagement of multiple stakeholders to identify the path towards feasible solutions. Think of the UN Sustainable Development Goals (SDGs, https://sdgs.un.org), which have identified the direction to follow "for peace and prosperity for people and the planet, now and into the future". To address any of the targets identified by the 17 goals, careful weighing of past and novel evidence is required for arriving at effective and efficient decisions that can change the way in which we live. In such a context, expert knowledge is key as decisions need to be taken by integrating multiple information sources, incorporating accumulated experience and weighing uncertainty. At the same time, the amount of available evidence is growing exponentially. For instance, during the first year of the COVID-19 pandemic, more than 400.000 papers and preprints have been published on that topic.[2] More generally, the total number of scientific publications is estimated to double every 14 years [1]. Such a rate of knowledge production makes it impossible for anybody to keep up to date, making it difficult to provide evidence-based interventions. Moreover, such situations run the risk of incurring into confirmation bias or cherry-picking. In a rapidly globalising world, solutions to complex problems are beyond the reach of isolated individuals. We need systems that can tap into the collective intelligence of multiple experts, that can make sense of massive amount of data and that aggregate different solution sources effectively, efficiently, and transparently.

A promising way to improve decision making in complex problems is exploiting collective intelligence (CI), which integrates the advice of multiple experts providing decision support [2,3,4,5]. While the power of CI has been successfully demonstrated in multiple domains, it was often applied to numerical predictions or binary-choice problems [6,7,8,9]. Complex problems are instead "open-ended", meaning that the solutions are not constrained to a (predefined, limited) set of alternatives. Instead, open-ended problems present a large, possibly infinite set of solutions, which could be conceptually different and not amenable to simple aggregation methods like, for instance, averaging a numerical estimate. Hence, extending CI to open-ended problems requires the ability to properly manipulate and combine the knowledge provided by multiple experts, a step that requires advanced domain-specific AI and data-based solutions.

Within the EU funded project HACID (http://www.hacid-project.eu), we propose a hybrid CI of human experts and AI systems [10]. Our goal is to build a decision support system (DSS) capable of (i) providing support for evidence-based decision making, and (ii) aggregating and expanding the solutions provided by experts, ultimately providing higher efficacy (e.g., increased accuracy of the solutions) and efficiency (e.g., reduced costs in terms of time or energies required), as well as higher user satisfaction, explainability and trust. The proposed system leverages complementarities between domain expertise from humans and the AI ability of reasoning and analysing vast amounts of data. In this way, we want to develop a general methodology to address complex, high stakes application domains. We illustrate our approach by applying it to two SDG-related domains: general medical diagnostics and climate change adaptation and risk management.

Medical diagnostics entails decision-making processes that can tell the difference between life and death. It is also a domain that sees large inequalities between developed

---

[2]Estimation from Google Scholar by counting 2020 results that have the keyword "COVID-19" in the title.

and developing countries. In medical diagnostics, the identification of a disease from a set of symptoms may be particularly complex, due to subject-specific conditions and comorbidities. As a matter of fact, the decision-making process is often open-ended as it deals with a large variety of possible diseases. Such open-endedness can exacerbate diagnostic errors both in the definition and interpretation of diagnoses, and such errors can have a rife of consequences, such as death, delayed treatments, but also an erosion of trust in the health system (e.g., when diagnoses from different physicians are not aligned or contradictory) [11,12]. Our tenet is that hybrid CI can improve the effectiveness and efficiency of diagnostic processes. Additionally, the human-centred approach and the explainable diagnostic solutions can contribute to generating more trust into AI-based technologies for healthcare.

The second open-ended domain we will highlight concerns climate change adaptation and risk management. Compared to medical diagnostics, it is a relatively new area of decision-making but also already supported by large formal and informal bodies of knowledge. Owing to the uncertainty in long-term climate projections, as well as the challenges of mapping climate model simulations to local-scale impacts [13], determining robust adaptation strategies for local contexts is difficult. To ensure that the cities we live in are resilient to future climate changes, evidence-based policy making is essential and is currently driving the nascent area of climate services [14]. Assessing and selecting interventions that are robust to future changes requires integrating multiple knowledge domains (climate, environment, human behaviour and social sciences as well as engineering, risk management and decision theory) in local contexts. The scientific evidence is vast and procedural standards for using the evidence to support decisions are only starting to be defined (EN ISO 14090:2019). Indeed, combining the evidence and uncertainty in projecting environmental and societal futures is particularly challenging. The existing information is often not salient, credible and legitimate [15] making adaptation planning intractable. Hybrid CI can help in aggregating the multiple disciplinary perspectives into a coherent picture, identifying relevant evidence, integrating climate data and treating uncertainty.

In this paper, we first present the concept behind the development of a hybrid CI for decision support, and provide the relevant background beyond which progress can be made (see Section 2). Section 3 introduces the medical diagnostics use case along with preliminary results, which illustrate the potential of the proposed approach. Section 4 presents a possible implementation in the climate services context. Finally, Section 5 concludes the paper with an outlook on hybrid CI research.

2. Concept and Background

The process of creating a hybrid CI that can exploit experts' and machine-generated knowledge can be synthesised into the following steps, as depicted in Figure 1: (A) knowledge engineering, to create and update a domain-specific knowledge base that supports hybrid CI; (B) knowledge refinement, to highlight those parts of the knowledge base that are relevant for a specific case; (C) hybrid collective problem solving, to elicit solutions from experts and aggregate them in a collective solution; and (D) evaluation, to determine effectiveness, efficiency and trustworthiness of the solution. In the following, we discuss each step and the challenges it entails.
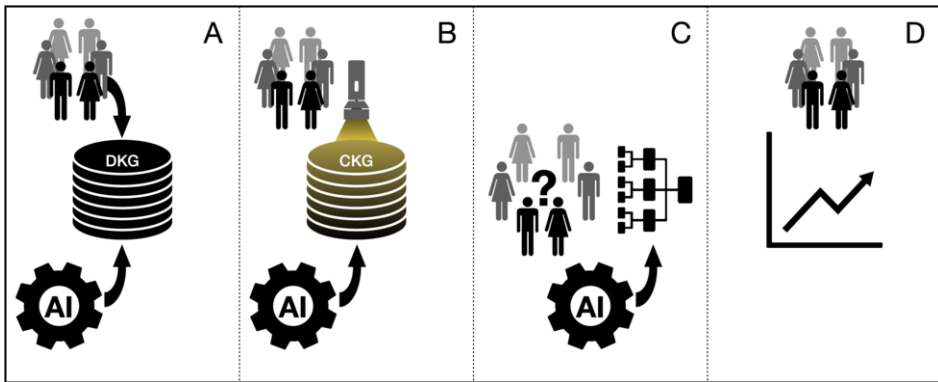
Figure 1. The four steps informing hybrid CI for decision support. (A) A domain knowledge graph (DKG) is created by integrating knowledge engineering methods to grasp expert knowledge and AI-generated knowledge obtained from the analysis of large bodies of evidence (e.g., scientific literature). (B) For every specific case (e.g., a patient with given symptoms), a case knowledge graph (CKG) is generated by collective tagging of relevant concepts as well as by knowledge enriching by both human experts and AI systems. (C) Solutions to the given case are provided by domain experts, and aggregated by reasoning on the concepts of the CKG or expanded, hence proposing novel solutions. (D) Evaluation of the solution is performed with respect to efficacy, efficiency and trustworthiness exploiting a participatory evaluation approach.

2.1. Knowledge engineering

Knowledge within a given domain is typically scattered. To make such knowledge available and usable, it needs to be retrieved and organised into a wide, structured knowledge base (a **domain knowledge graph**, **DKG**) that can be explored in search of evidence. Knowledge Graphs (KGs) are becoming increasingly popular with both big industrial players and scientific communities in different research fields such as Semantic Web, Databases, Machine Learning and Data Mining. Though there is no consensus on the exact definition of KGs [16], here we refer to linked open data including both schema axioms (i.e., Web Ontology Language, OWL) and factual data (i.e., Resource Description Framework, RDF triples). An increasing number of methodologies is becoming available for the construction of KGs with most of them focusing on maximising the discovery and reuse of symbolic knowledge, as dictated by the FAIR (Findable, Accessible, Interoperable, Reusable) principles. More recently, agile methodologies, such as the eXtreme Design [17,18], proved to be very effective for the construction of large KGs [19,20,21]. However, knowledge can be extracted from a plethora of different data sources, which are inherently heterogeneous with respect to syntax and semantics, i.e., the *knowledge soup* problem [22]. Additionally, in many cases the data sources contain billions of datapoints that need to be processed. It is difficult to synthesise all the available scientific and technical knowledge required to inform decision making—especially for complex issues like medicine and climate change—leading to the so-called *information horizon* problem [22]. Facing both the knowledge soup and information horizon problems is challenging for creating KGs at scale. Recent advances in open knowledge extraction [23,24] exploit *machine reading* [25] as a sub-symbolic paradigm for gathering structured knowledge from text by relying on natural language processing (NLP) and deep learning. A DKG can be therefore generated by harmonising symbolic and sub-symbolic knowledge, in close collaboration with domain experts, stakeholders and end-users. This

step is performed for each specific domain to gather all the relevant knowledge for answering specific questions (i.e., specific *cases*). The DKG is not static information, but evolves by aggregating new evidence as soon as this is produced and validated by the relevant stakeholders. It is both flexible and customised.

2.2. Case knowledge refinement

There are clear limits to how well humans can process large amounts of information [26,27]. Vast domain knowledge—even if well structured in a KG—can lead to cognitive overload if information is not properly presented for a specific case to be solved, possibly leading to sub-optimal knowledge extraction through, for example, confirmation bias or cherry picking. Or the complexity of the domain knowledge might challenge the perceptual and cognitive limits of experts. Even AI approaches may struggle with large knowledge bases if the solution landscape does not clearly provide preferential search directions [28]. In both cases, pruning the knowledge space before exploring it can lead to knowledge refinement and better chances of identifying the correct solution or identifying it more efficiently (see below). To this purpose, advanced knowledge exploration and visualisation tools offer a great range of possibilities to condense, simplify, and make sense of large data sets. These tools can support cognitive processes humans use to search for information outside or inside their mind [29,30,31]. However there are also limitations associated with existing visualisation and exploration tools. Many dashboards use a fixed number of features to communicate a complex situation, with the associated risk of providing an incomplete description of most real-world tasks. Simplified representations can influence decision making by eliminating nuances, sometimes leading to worse outcomes [32]. In addition, dashboards oriented to visualisation are often restricted to display available information, and do not allow opportunities for commenting, discussion, or other features that can support the production of further information by expert users. To address complex open-ended domains, new approaches are needed that can manage large knowledge bases where the size and scope of the relevant evidence is not known a priori [33].

Next to supporting exploration using suitable interfaces, there is also the potential to improve how domain knowledge is navigated using tools from network science, which offers many approaches for measuring the relevance or importance of nodes in a graph [34]. However, there is a lack of methodologies to apply these network approaches to knowledge engineering and knowledge discovery, in particular, for determining what parts of the knowledge base are worth considering and what needs to be pruned.

Advanced knowledge exploration and visualisation tools and computational methodologies informed by cognitive network science [35,36,37] can help users identify relevant sources of evidence for a specific case within a domain, thus facilitating hybrid human and machine problem-solving and decision making downstream. Given a particular case, experts are asked to provide relevant information that can support decision making. On the basis of the responses provided, the DKG is refined with that case-specific knowledge forming a **case knowledge graph** (**CKG**). The CKG is not just a subset of the DKG, it is rather the combination of the DKG and a collective annotation layer that enriches the DKG with case-specific information. The relevance/importance of the nodes in the CKG can be quantified using a network node centrality measure suitable for the domain at hand [34]. Importantly, in this graph, even nodes that no individual expert

nominated can be central if they are strongly connected to many nominated nodes. This is the first outcome of a hybrid CI, because the input of multiple experts can be integrated for the creation of knowledge useful to provide structured background information. The AI part is not limited to the processing and aggregation of human-generated evidence. It can also directly participate in the information extraction by analysing the available case information, hence selecting relevant knowledge from the DKG autonomously and independently from expert input.

2.3. Hybrid collective problem solving

CI is the ability of groups of individuals to solve problems exceeding the capacities of single individuals [3,4,5,2]. Previous work has shown the tremendous potential of CI to boost decision accuracy for problems with a well-defined and finite set of answer options (e.g., simple categorisation tasks, quantitative estimates, subjective probability distributions), in diverse domains, such as geopolitical forecasting, lie detection, cancer diagnostics and fingerprint recognition [38,8,39,40,41]. However, applying CI to open-ended domains has seen only limited applicability, and approaches that can harness the power of the collective in such domains are largely lacking. This is an important gap, as our increasingly complex world requires us to make decisions in situations where there is high uncertainty both about the relevant information sources and the outcome space.

A tool based on hybrid CI would harness the expertise from a diverse set of experts to propose solutions that surpass the performance of the best expert within the group, and possibly expand the solution space beyond what was initially provided by individual experts. To this end, human experts are asked to provide a set of possible solutions to a given case. These solutions can be in the form of open-ended text, and interaction with the CKG can guide the user in producing them. Additional metadata can be collected as well, for instance related to confidence or competence of the user with respect to the specific case [42,7]. Also, interaction among users can be enabled, to profit from social feedback in the elicitation of responses [43,44]. The user input is then processed by the underlying AI system that aggregates and enriches the available solutions to create a *collective solution*. This means recognising the (dis)similarity in the provided solutions (e.g., identifying synonyms), aggregating them into classes and providing a unified view. Possibly, the set of solutions can be expanded with solutions that were not initially provided by any expert. The AI will also "explain" the collective solution, linking to the relevant evidence from the CKG and demonstrating the steps that lead to aggregation and expansion of the experts' input.

2.4. Evaluation

While there are multiple methodologies for evaluating the accuracy of human or machine decisions about problems with a well-defined or finite set of answer options (e.g., simple categorisation tasks, quantitative estimates, subjective probabilities [45,46,47,48,49,50,51,52]), methods to rigorously evaluate solutions to open-ended problems are largely lacking. Furthermore, in many important use cases where a DSS is sought to improve decision making, there is no ground truth against which accuracy can be assessed during development of the tool and/or monitoring it after deployment. This may be because the ground truth will only be available in the (long-term) future

(i.e., forecasting) and/or because the relevant ground truth is not available in principle (e.g., in case of a self-defeating prophecy where, for example, a prediction turns out to be false because a player intervenes to avoid the negative outcome). In similar cases, expert evaluation is necessary to determine the relevance of proposed solutions, and to evaluate other aspects like diversity of options, representativeness of minorities and gender balance. We maintain that a participatory approach to the development and design of the DSS (broadly referred to under the umbrella term "Participatory AI") can help determine what features could facilitate higher levels of understanding, trust and confidence for intended end-users and other stakeholders. Measuring the performance of the DSS against these criteria represent an invaluable method for comprehensive evaluation.

3. Crowd-sourcing medical diagnostics

The hybrid CI approach discussed above can be deployed for open-ended problems in different domains. We discuss here the use case of general medical diagnostics, providing also preliminary results of hybrid CI in context.

This use case leverages the online platform developed by the Human Diagnosis Project (Human Dx, https://www.humandx.org), which is open to medical practitioners who can post cases and crowdsource input from a large pool of professionals worldwide. Each medical case is associated with a short description of the patient and several insights (symptoms, results of medical tests). The system then prompts the users to provide their independent diagnoses, ranked according to what they believe corresponds best to the case.

A hybrid CI approach to medical diagnostics can expand the current platform by (i) defining a DKG that enables a rich description of the domain, including medical terms, prevalence information, treatments and so forth; (ii) allowing experts to provide evidence in support of each case to create a CKG; (iii) enhancing the user interface allowing to introduce subjective confidence and competence estimates, as well as enabling social interactions among users; and (iv) deploying methods for reasoning about the user inputs over the CKG. Eventually, the system summarises all the answers provided by the expert users into a collective diagnosis, giving more weight to concepts that are supported by large parts of the user crowd, and possibly expanding the list with other related concepts from the CKG.

3.1. A Preliminary Approach to Medical Diagnostics

In a previous study, we took the first steps in the definition of a hybrid CI for medical diagnostics, deploying a preliminary version of the DSS and evaluating it on a sample dataset provided by Human Dx [53]. Here, we present previous results and we improve by optimising the aggregation method. The dataset consists of cases with known diagnoses, allowing us to make a ground-truth comparison. In order to identify and reason on the user-provided diagnoses, we use a DKG based on SNOMED-CT, an open, structured collection of clinical terms [54,55]. The differential diagnoses introduced by users about a case—hereafter referred to as solutions—are aligned with SNOMED-CT terms, enabling operation on unique concepts. Challenges here are related to the many different ways in which users refer to the same concept, using acronyms or entire sentences.

Hence, a matching problem must be solved, that is, every diagnosis provided by a user must be associated with a specific formal concept that represents the disease that the users wanted to propose. Second, it is necessary to compare different diseases with each other, to understand whether they are equivalent and also whether or not a proposed diagnosis is correct with respect to the available ground truth. The equivalence between two diagnoses is not straightforward, because of possible referral to sub- or super-classes of a disease (e.g., *Aortic aneurysm* is a more generic term than *Abdominal aortic aneurysm*), and the substantial freedom of users to provide solutions that may refer not just to the disease, but to the causes (e.g., specifying a virus instead of the disease it causes).

The matching problem has been addressed by computing the Jaccard similarity between solutions (after a necessary normalisation step using routine NLP methods) and the available descriptions of the SNOMED-CT concepts. Out of the initial 1572 cases available, we obtained 1333 cases in which the available ground truth could be exactly matched (Jaccard similarity of 1) [53]. Then, we processed the solutions provided by the users to compute a collective differential diagnosis, also attempting to match them against SNOMED-CT concepts with maximum Jaccard similarity. For each case, we performed the following steps:

- Randomly draw a group of N users among those that solved the given case.
- Retain the best N_u solutions for each user, and weigh each solution by a factor $R(n)$, where n is the original rank of the solution in the user-provided list.
- Aggregate solutions from all users into a ranked list by summing up the scores of matching SNOMED-CT concepts.
- Check whether the correct diagnosis is in the top N_c positions, exploiting a function $M(p,s)$ that determines if a solution s actually matches the ground truth p.

3.2. Results

The obtained results are clearly dependent on the functions $R(n)$ and $M(p,s)$ used for computation of the results. While previous studies propose an inversely proportional relation between score and rank (i.e., $R_i(n) = n^{-1}$) [56,53], in this paper we search for optimal values by means of a brute-force approach, which we evaluated only for the case $N = 10$ and $N_u = N_c = 5$, allowing weights to vary in the interval $[0,1]$. Also, differently from [53], here we include all equally ranked solutions as long as the rank does not exceed the threshold N_c (e.g., if there are three solutions with rank equal to $N_c = 1$, these
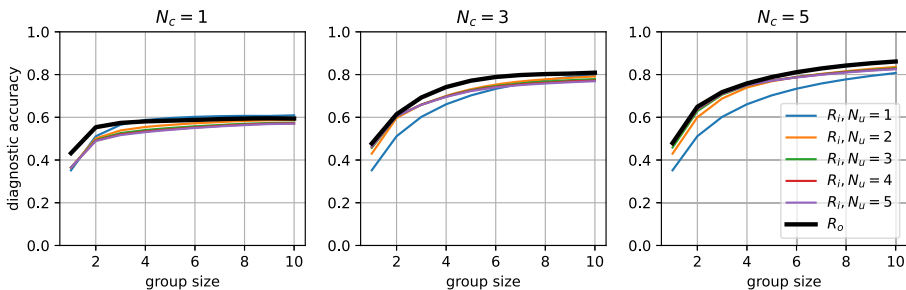

Figure 2. Diagnostic accuracy for a perfect matching between solutions and ground truth
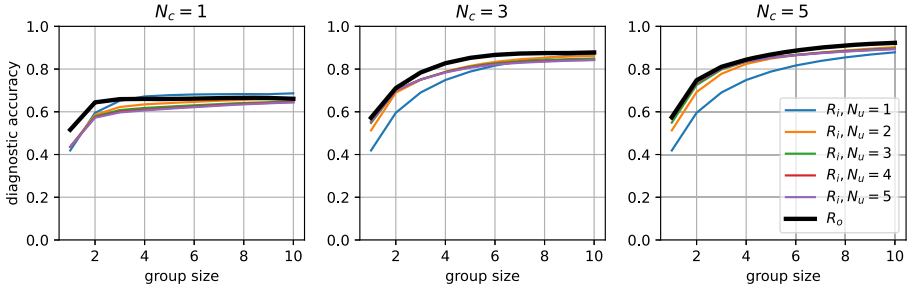
Figure 3. Diagnostic accuracy exploiting the hierarchical organisation of medical concepts in SNOMED-CT

are all considered). We have found that the best scoring in terms of average diagnostic accuracy equally considers the first two ranked diagnoses, and completely disregards the other ones: $R_o(n) = 1$ if $n \in \{1, 2\}$, otherwise $R_o(n) = 0$. Figure 2 shows the results when a perfect matching between solution and ground truth is considered (i.e., $M(p, s) = 1$ if $p = s$). In such conditions, the maximum diagnostic accuracy reaches 86% (as opposed to 48% for individual raters), and remains highest for every group size, provided that N_c is rather high. Lower values of N_c would likely require a different ranking $R(n)$, as the current one has been optimised for $N_c = 5$.

Even better results can be obtained if we exploit the structure of SNOMED-CT—which is organised as a poly-hierarchy—to determine whether a proposed diagnosis is acceptable with respect to the available ground truth. Specifically, we consider that a diagnosis is correct when it is (i) equal to the ground truth; (ii) a subclass of the ground truth (i.e., the diagnosis is more specific than the ground truth); (iii) the direct superclass of the ground truth (i.e., the distance between the two concepts is minimal); or (iv) a sibling of the ground truth, that is, it shares the same superclass (i.e., the two concepts have a single common ancestor). In this case, many more solutions match the ground truth, and the accuracy generally increases for all the tested methods. Figure 3 shows that the maximum accuracy increases up to 92.3% (against 57.5% of individual accuracy).

These preliminary results confirm the suitability of a collective intelligence approach to open-ended diagnostic problems, considering that the group accuracy significantly increases over individual performance. Further developments and evaluation of the proposed methods are therefore very promising.

4. A Conceptual Framework for Applying Hybrid CI to Climate Services

This use case focuses on improving existing climate services that support decision-makers in adapting and managing their cities/country to improve resilience to uncertain future climate changes. The DSS can be applied to many contexts and for multiple climate hazards. For instance, climate service providers are currently serving UK stakeholders including members of the London Climate Change Partnership (http://climatelondon.org/) who are preparing their city for extreme weather today and climate change in the future. Here, we discuss a conceptual framework for hybrid CI in climate services based on ongoing user research.

First and foremost, we identify the object of the climate-related decision process. Generally speaking, climate change adaptation management usually requires the assessment of climate risks, which includes information on the hazard, exposure and vulnerability for a given location and time-frame (e.g., for the Thames Estuary 2100 plan, adaptation was considered for the Greater London area for a 100-year timespan [57]). Hence, the DKG is populated with information about past climate and climate model projections (e.g., including those that fed into the latest 6th Assessment Report from the Intergovernmental Panel on Climate Change—i.e., the Climate Model Intercomparison Project Phase 6 [58]—as well as information supported by National Governments—e.g., UK [59], Switzerland [60], The Netherlands [61]). We also include approaches for the sub-selection (e.g., merging, selecting and/or filtering) of climate models and associated information. Relationships between models are captured (e.g., models from the same model family with similar components or assumptions [62]) as well as their suitability for different adaptation management decision strategies [63]. Finally, we consider domain-specific information (i.e., exposure and vulnerability) related to the implementation into an adaptation management strategy and the decisions that need to be made. This latter aspect is crucial to customise the DSS for specific decision domains, as the level of information required is likely different. For instance, risk assessment for urban adaptation strategies requires information about urban plans, land-use, infrastructure and socio-economics.

Given a specific case (e.g., increased surface water flood risk in London potentially caused by increases in extreme rainfall intensity), experts are first asked to identify, within the DKG, the information that they deem relevant. For instance, selected information can be related to relevant metrics at a suitable temporal and spatial scale that need to be projected to support the impact models and feed into adaptation options evaluation. Besides, experts can identify in the DKG or provide further information relevant for the case, including scientific and grey literature applying alternative methods and information sources, or previous case studies. The results from the experts' input is the CKG on which reasoning will be made.

The collective problem solving step requires experts to provide their own answers to the selection of appropriate climate information, from diverse perspectives. We initially consider answers in terms of identifying and ranking the climate information to be used, and possibly combining this with aggregation methods. Indeed, there is no strong agreement among experts about which set of climate information would best fit a given use case, and there are also many methods to combine (or not) climate projections to explore uncertainty and build a coherent picture of the range of possible futures. Hence, the combination of the solutions from multiple experts can provide suggestions for suitable identification and aggregation approaches of the most relevant models for the use case. In a second step, we consider more freedom in the proposed solution, gathering additional information from a heterogeneous set of experts representing those closer to the decision and policy-making end of the solution space. This means that even the relevant variables and their expected effects are targeted by the collective decision process, generating a rich description of the possible solutions to the case that will contribute to the generation of a case report in support of policy making.

Within the HACID project, we aim at evaluating the collective solution in comparison with an individually-based solution according to internationally-accepted guidelines. For instance, reference cases can be derived from the studies considered in the EUCP

project (https://www.eucp-project.eu) [64] that use state-of-the-art high-resolution climate models to calculate flood risk across multiple European cities. Also, OpenCLIM (http://climatereadyclyde.org.uk/openclim/) provided several case studies focused on Glasgow, UK, which feature a community modelling approach. These case studies can provide both a reference and relevant baseline for comparison. Additionally, a user-centred approach will be also considered to understand acceptability and trustworthiness of the DSS. This is framed in a participatory AI approach that informs the technology development from day one, involving stakeholders in the definition of every component of the DSS largely before the final evaluation step. We hope in this way to capture the real needs and expectations of the community revolving around climate change adaptation and risk management, delivering a tool that really responds to their needs.

5. Discussion and Conclusions

We presented a framework for the exploitation of hybrid CI in context, and preliminary results supporting the proposed approach within the medical diagnostics use case. These results still need to be accurately validated, especially when the hierarchy of concepts in SNOMED-CT is exploited for evaluation. However, they indicate that a hybrid CI approach can significantly improve over individual diagnosticians.

A key aspect of the proposed approach is that it can be rooted in specific domains, such as the medical diagnostics and climate services we presented here, which testifies to the large potential impact achievable by hybrid CI. With respect to medical diagnostics, it is fair to admit that AI is conquering everyday new grounds, for instance in supporting radiologist in spotting the most urgent cases. Very often, AI-based diagnostic systems are tailored to very specific diseases and are based on some image classification software. However, actual deployment is very limited due to lack of trust and medical ethics issues [65]. The hybrid CI approach can have a concrete impact because it addresses a wider domain than current AI technologies, and because it is designed to exploit the complementary abilities of human experts and AI systems, hence improving trust in the system.

Climate services for supporting climate change adaptation and risk management currently lack technologies that can support navigating an already very large evidence base. The potential impact of hybrid CI is therefore very large, considering that a competing service does not exist to date. National and international agencies would benefit from a more structured, unbiased and evidence-based decision process. We believe that different cities—e.g., within the C40 Cities Climate Leadership Group (https://www.c40.org)—would be interested in testing the DSS for adaptation management, comparing it with current practices.

Other application domains may have a large, still unexplored potential for the application of hybrid CI. The essential features that qualify a domain to be amenable to the proposed hybrid CI technology are: (i) a large, diverse and possibly unstructured evidence base, which leads to the need of formalising the domain knowledge in a DKG; (ii) recurrent cases with diverse features that cannot be equated to the same decision problem (i.e., are not amenable to a standard supervised ML approach), which leads to the need of determining an always changing CKG; (iii) a pool of users/experts possibly with different backgrounds and expertise, which lead to the need of integrating diverse

opinions into a collective outcome; (iv) open-endedness of the possible solutions to a given case, which leads to the need of finding similarities and relatedness among the possible answers. We firmly believe that, when these characteristics are found, hybrid CI can truly shine.

Acknowledgements

References

[1] Bornmann L, Haunschild R, Mutz R. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanities and Social Sciences Communications. 2021;8(1):224.

[2] Surowiecki J. The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. Economies, Societies and Nations. 2004;296(10.5555):1095645.

[3] Krause J, Ruxton GD, Krause S. Swarm intelligence in animals and humans. Trends in ecology & evolution. 2010;25(1):28-34.

[4] Malone TW, Laubacher R, Dellarocas C. The collective intelligence genome. MIT Sloan management review. 2010;51(3):21.

[5] Suran S, Pattanaik V, Draheim D. Frameworks for collective intelligence: a systematic literature review. ACM Computing Surveys. 2020;53(1):1-36.

[6] Mannes AE, Soll JB, Larrick RP. The wisdom of select crowds. Journal of Personality and Social Psychology. 2014;107(2):276-99.

[7] Budescu DV, Chen E. Identifying expertise to extract the wisdom of crowds. Management Science. 2015;61(2):267-80.

[8] Herzog SM, Litvinova A, Yahosseini KS, Tump AN, Kurvers RH. The Ecological Rationality of the Wisdom of Crowds. In: Hertwig R, Pleskac TJ, Pachur T, The Center for Adaptive Rationality, editors. Taming uncertainty. MIT Press; 2019. p. 245-62.

[9] Marshall JA, Kurvers RH, Krause J, Wolf M. Quorums enable optimal pooling of independent judgements in biological systems. Elife. 2019;8:e40368.

[10] Pescetelli N. A brief taxonomy of hybrid intelligence. Forecasting. 2021;3(3):633-43.

[11] Donaldson MS, Corrigan JM, Kohn LT, et al. To err is human: building a safer health system. National Academies Press; 2000.

[12] Makary MA, Daniel M. Medical error—the third leading cause of death in the US. British Medical Journal. 2016;353.

[13] Lemos MC, Kirchhoff CJ, Ramprasad V. Narrowing the climate information usability gap. Nature Climate Change. 2012;2:789-94.

[14] WMO Secretariat. WMO Supporting the 2030 Agenda for Sustainable Development; 2016. Bulletin n.: Vol 65(1).

[15] Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, et al. Knowledge systems for sustainable development. Proceedings of the National Academy of Sciences. 2003;100(14):8086-91.

[16] Bonatti PA, Decker S, Polleres A, Presutti V. Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Dagstuhl Reports. 2019;8(9):29-111.

[17] Presutti V, Daga E, Gangemi A, Blomqvist E. eXtreme Design with Content Ontology Design Patterns. In: Blomqvist E, Sandkuhl K, Scharffe F, Svátek V, editors. Workshop on Ontology Patterns. vol. 516 of CEUR Workshop Proceedings. CEUR-WS.org; 2009. p. 83–97.

[18] Blomqvist E, Hammar K, Presutti V. Engineering Ontologies with Patterns - The eXtreme Design Methodology. In: Hitzler P, Gangemi A, Janowicz K, Krisnadhi A, Presutti V, editors. Ontology Engineering with Ontology Design Patterns. vol. 25 of Studies on the Semantic Web. IOS Press; 2016. p. 23-50.

[19] Presutti V, Lodi G, Nuzzolese A, Gangemi A, Peroni S, Asprino L. The Role of Ontology Design Patterns in Linked Data Projects. In: Comyn-Wattiau I, Tanaka K, Song IY, Yamamoto S, Saeki M, editors. Conceptual Modeling. vol. 9974 of Lecture Notes in Computer Science. Cham: Springer; 2016. p. 113-21.

[20] Nuzzolese AG, Gentile AL, Presutti V, Gangemi A. Conference Linked Data: The ScholarlyData Project. In: Groth P, Simperl E, Gray AJG, Sabou M, Krötzsch M, Lécué F, et al., editors. The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part II. vol. 9982 of Lecture Notes in Computer Science. Cham: Springer; 2016. p. 150-8.

[21] Carriero VA, Gangemi A, Mancinelli ML, Nuzzolese AG, Presutti V, Veninata C. Pattern-based design applied to cultural heritage knowledge graphs. Semantic Web. 2021;12(2):313-57.

[22] Gangemi A, Presutti V. Towards a pattern science for the Semantic Web. Semantic Web. 2010;1(1-2):61-8.

[23] Gangemi A, Presutti V, Recupero DR, Nuzzolese AG, Draicchio F, Mongiovì M. Semantic Web Machine Reading with FRED. Semantic Web. 2017;8(6):873-93.

[24] Presutti V, Nuzzolese AG, Consoli S, Gangemi A, Recupero DR. From hyperlinks to Semantic Web properties using Open Knowledge Extraction. Semantic Web. 2016;7(4):351-78.

[25] Etzioni O, Banko M, Cafarella MJ. Machine Reading. In: AAAI Spring Symposium: Machine Reading. AAAI; 2007. p. 1-5.

[26] Hills TT. The dark side of information proliferation. Perspectives on Psychological Science. 2019;14(3):323-30.

[27] Simon HA. Designing organizations for an information-rich world. In: Greenberge M, editor. Computers, communications, and the public interest. Baltimore, MD: Johns Hopkins Press; 1971. p. 37-52.

[28] Russell SJ, Norvig P. Artificial intelligence: A modern approach. 4th ed. Harlow: Pearson; 2022.

[29] Fu WT, van Oostendorp H. Understanding and improving information search: A cognitive approach. Cham: Springer; 2020.

[30] Hills TT, Todd PM, Lazer D, Redish AD, Couzin ID. Exploration versus exploitation in space, mind, and society. Trends in Cognitive Sciences. 2015;19(1):46-54.

[31] Todd PM, Hills TT. Foraging in mind. Current Directions in Psychological Science. 2020;29(3):309-15.

[32] Singh M. Heuristics in the delivery room. Science. 2021;374(6565):324-9.

[33] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:170208608. 2017.

[34] Newman M. Networks. vol. 1. Oxford University Press; 2018.

[35] Herzog SM, Hills TT. Mediation centrality in adversarial policy networks. Complexity. 2019;2019:1-15.

[36] Hills TT, Kenett YN. Is the mind a network? Maps, vehicles, and skyhooks in cognitive network science. Topics in Cognitive Science. 2021;14:189-208.

[37] Kenett YN, Beckage NM, Siew CSQ, Wulff DU. Cognitive network science: A new frontier. Complexity. 2020;2020:1-4.

[38] Arrow KJ, Forsythe R, Gorham M, Hahn R, Hanson R, Ledyard JO, et al. The promise of prediction markets. Science. 2008;320(5878):877-8.

[39] Klein N, Epley N. Group discussion improves lie detection. Proceedings of the National Academy of Sciences. 2015;112(24):7460-5.

[40] Kurvers RH, Herzog SM, Hertwig R, Krause J, Carney PA, Bogart A, et al. Boosting medical diagnostics by pooling independent judgments. Proceedings of the National Academy of Sciences. 2016;113(31):8777-82.

[41] Mellers B, Ungar L, Baron J, Ramos J, Gurcay B, Fincher K, et al. Psychological strategies for winning a geopolitical forecasting tournament. Psychological science. 2014;25(5):1106-15.

[42] Marshall JAR, Brown G, Radford AN. Individual Confidence-Weighting and Group Decision-Making. Trends in Ecology & Evolution. 2017;32(9):636-45.

[43] Bang D, Frith CD. Making better decisions in groups. Royal Society Open Science. 2017;4(8):170193

22.

[44] Becker J, Brackbill D, Centola D. Network dynamics of social influence in the wisdom of crowds. Proceedings of the National Academy of Sciences. 2017;114(26):E5070 E5076.

[45] Brown NR, Siegler RS. Metrics and mappings: A framework for understanding real-world quantitative estimation. Psychological Review. 1993;100(3):511-34.

[46] Carvalho A. An overview of applications of proper scoring rules. Decision Analysis. 2016;13(4):223-42.

[47] Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A, Pocernich M, et al. Forecast verification: Current status and future directions. Meteorological Applications. 2008;15(1):3-18.

[48] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27(8):861-74.

[49] Hand DJ. Assessing the performance of classification methods. International Statistical Review. 2012;80(3):400-14.

[50] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006;22(4):679-88.

[51] Macmillan NA, Creelman CD. Detection theory: a user's guide. Mahwah, NJ: Lawrence Erlbaum Associates; 2005.

[52] O'Hagan A, Buck CE, Daneshkhah A. Uncertain judgements: Eliciting experts' probabilities. Chichester: John Wiley & Sons; 2006.

[53] Kurvers RHJM, Nuzzolese AG, Russo A, Barabucci G, Herzog SM, VT. Automating Hybrid Collective Intelligence in Open-Ended Medical Diagnostics. PsyArXiv. 2023:1-8.

[54] Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. In: Proceedings of the AMIA annual fall symposium. American Medical Informatics Association; 1997. p. 640.

[55] Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Studies in health technology and informatics. 2006;121:279.

[56] Barnett ML, Boddupalli D, Nundy S, Bates DW. Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians. JAMA Network Open. 2019;2(3):e190096.

[57] Ranger N, Reeder T, Lowe JA. Addressing deep uncertainty over long-term climate in major infrastructure projects: four innovations of the Thames Estuary 2100 Project. EURO Journal on Decision Processes. 2013;11:233-62.

[58] Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development. 2016;9(5):1937-58.

[59] Lowe JA, Bernie D, Bett P, Bricheno1 L, Brown S, Calvert D, et al. UKCP18 Science Overview Report. Met Office; 2018.

[60] CH2018. CH2018 – Climate Scenarios for Switzerland. National Centre for Climate Services; 2018.

[61] Attema J, Bakker A, Beersma J, Bessembinder J, Boers R, Brandsma T, et al. KNMI'14: Climate Change scenarios for the 21st Century—A Netherlands perspective. KNMI; 2014.

[62] Brunner L, Lorenz R, Zumwald M, Knutti R. Quantifying uncertainty in European climate projections using combined performance-independence weighting. Environmental Research Letters. 2019;14(12):124010.

[63] Siders AR, Pierce AL. Deciding how to make climate change adaptation decisions. Current Opinion in Environmental Sustainability. 2021;52:1-8.

[64] Zander MJ, Sperna Weiland FC, Weerts AH. Introducing H2020 European climate prediction system for urban pluvial floods. In: UrbanRain18: 11th International Workshop on Precipitation in Urban Areas; 2018. .

[65] Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthcare Journal. 2019;6(2):94-8.