chist-era

CHIST-ERA Project Periodic Report

EXPECTATION

PERSONALIZED EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR DECENTRALIZED AGENTS WITH HETEROGENEOUS KNOWLEDGE

Periodic report n°2 [24M]
[1/05/2023]

1. Progress Report

1.1. Project objectives and activities implemented *(Indicative length: 2 pages per period)*

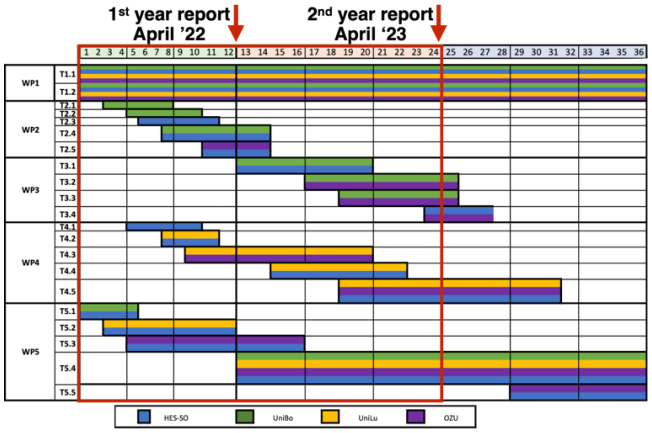

Figure 1.1 – EXPECTATION's Gantt.

Figure 1 shows the EXPECTATION's Gantt chart, indicating with a color code the associations tasks -- partner. The HES-SO team is led by Prof. Michael Schumacher and composed of Dr. Davide Calvaresi (senior researcher – main expertise on distributed intelligent systems, real-time systems, and explainable AI) and Victor Contreras (Ph.D. student – main expertise on Machine Learning and medical imaging) since M1. The OZU team is led by Dr. Reyhan Aydoğan and composed of her graduate students (Furkan Cantürk, Berk Buzcu, and Anıl Doğru) and her bachelor student (Berkecan Koçyiğit) at Özyeğin University. The UNIBO team is led by Prof. Andrea Omcini and composed of PhD students and fixed-term research fellows at University of Bologna. These include Giovanni Ciatto (Ph.D. student and WP leader since M1), Federico Sabbatini (who contributed during M1-8), A. Agiollo (PhD student, who contributed occasionally), and M. Magnini (research fellow, who contributed since M8). The UNILU team is led by Prof. Leon van der Torre and composed of three postdoctoral researchers named Dr. Amro Najjar (since M1), Dr. Joris Hulstijn (since M17) to work on social cues, and Dr. Igor Tchappi (since M10) expert on user-studies and subjective user test. Dr Joris is a senior research associate, with strong focus on agent dialogue. In addition to postdoc, the UNILU team involves also Rachele Carli, a PhD student at the UNILU and UniBO expert on legal AI, and Melissa Tessa an internship student from Ecole Nationale Supérieure d'Informatique (ESI) in Algeria who joined Luxembourg in March 2022 for a duration of 6 months. She focused on the development of the protocol, the robots and the integration of Large Language Models (LLMs). Finally, the last partner, LIST is mainly represented by Dr. Amro Najjar. The partners' contributions to the project are below organized per work package (WP).

chist-era

WP1

[T1.1 – M1-M36 (ongoing)] Project coordination:

All WP leaders (representing all the partners) have participated in the monthly meetings to steer the project. Each meeting, all the WP leaders or other members of their team have presented to the partners their advancements and contributions. Moreover, these latter have punctually been discussed generating further peer-to-peer follow-up collaborations/reviews.

[T1.2 – M1-M36 (ongoing)] Dissemination & Exploitation:

All the partners have worked to disseminate their achievements publishing several papers listed in Section 2.1, organizing scientific panels (see short description below (a)), and realizing a video for general-public dissemination (b).

(a) EXTRAAMAS 2021: Dr. Davide Calvaresi has led a scientific panel named "Distributed Intelligent Systems and XAI" hosted by the international workshop on Explainable and TRAnsparent AI and Multi-agent Systems (EXTRAAMAS), which was held in conjunction with AAMAS conference in May 2021 (https://extraamas.ehealth.hevs.ch/index.html) with the aim of promoting EXPECTATION. The speakers, core of the panel, have been Dr. Reyhan Aydoğan, Prof. Andrea Omicini, and Prof. Leon Van Der Torre.

(b) https://expectation.ehealth.hevs.ch/posts/promo/

(c) Visit to Senegal: The attitude towards AI in general, robots and nutrition virtual coaches is likely to be different between users from different cultural backgrounds. To investigate this research question, Igor Tchappi conducted a research stay in Senegal and started initial HRI experiments.

(d) EXTRAAMAS 2022: The International workshop of explainable and transparent agents was held successfully for the third year. Due to the covid restrictions, which were still in vigor at the time, the workshop was held online. For the first time, EXTRAMAS 2022 included a track on AI & Law. The track chair was Dr. Réka Markovich from the university of Luxembourg. Moreover, EXTRAAMAS features 2 keynote speakers. Bertram Malle (Brown University) with a presentation entitled “From Explanation to Justification to Trust in Human-Machine Interaction”, and Serena Vilatta (CNRS, Nice). The latter keynote was part of the AI & Law Track and was entitled: “Towards Natural Language Explanatory Argument Generation: Achieved Results and Open Challenges”. The acceptance rate of the workshop was around: 50%. The outcome was one Springer Volume with 16 accepted papers.

(e) Dissemination of the finding in one of the biggest worldwide conferences in AI named AAAI. One of the publications of UNILU, LIST, and HES-SO has been accepted in the Journal track of AAAI 2023 and was co-presented by Dr Davide Calvaresi and Igor Tchappi on the 6th of February 2023. In addition, a poster was designed and exhibited in the room dedicated to posters with many visitors, both researchers attending the conference and people passing by. UNILU, LIST and HES-SO explained the contributions of the paper to people asking for further details on the paper.

(f) Tutorial on PSyKE and DEXiRE: the HES-SO and UNIBO teams presented a tutorial session (3h) at the international conference on Principles and Practice of Multi-Agent Systems (PRIMA 2022) in Valencia, Spain. The tutorial is related to T2.1 and T2.4, where the UNIBO team has developed PSyKE knowledge extractor [wp1-1] and HES-SO team developed a library namely DEXiRE (Deep Explanation and Rule Extraction) to extract rule sets from deep learning predictors [wp1-2,wp1-3].

(g) CHIST-ERA Annual meeting (Bratislava): Dr. Davide Calvaresi has participated at the annual CHIST-ERA event for XAI-related funded projects. He has shared the consortium advancements creating links with other projects and planning the investigations about feasibility/realization of cross-projects joint XAI libraries by the end of the 3rd year.

References WP1

[wp1-1] Sabbatini, F., Ciatto, G., Calegari, R., & Omicini, A. (2021, September). On the Design of PSyKE: A Platform for Symbolic Knowledge Extraction. In *WOA* (pp. 29-48).

[wp1-2] Contreras, V., Marini, N., Fanda, L., Manzo, G., Mualla, Y., Calbimonte, J. P., ... & Calvaresi, D. (2022). A DEXiRE for extracting propositional rules from neural networks via binarization. Electronics, 11(24), 4171.
[wp1-3] Contreras, V., Bagante, A., Marini, N., Schumacher, M., Andrearczyk, V., & Calvaresi, D. (2023, May). Explanation Generation via Decompositional Rules Extraction for Head and Neck Cancer Classification. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems* (pp. 187-211). Cham: Springer Nature Switzerland.

WP2

[T2.1 – M3-M8 (completed)] Symbolic knowledge extraction:

The UNIBO team performed an extensive literature exploration which led to the identification of nearly 90 methods for symbolic knowledge extraction (SKE) and nearly 70 methods for symbolic knowledge injection (SKI). These were then unified and jointly categorized in a coherent framework, as part of the activities required by T2.2. The SKE/SKI investigation enabled the realization of T2.4, as well as deliverables D2.1, D2.2, and D2.3 (cf. Section 1.6 below).

[T2.2 – M4-M10 (completed)] Generalization:

The UNIBO team generalized the information gathered by a deep analysis of the methods for SKE and SKI surveyed in T2.1. Hence, they produced a coherent and unified theoretical (conceptual) framework. Both the framework and the surveyed methods are discussed in detail in D2.1, composing a technical report (nearly 110 pages long). This report is currently being synthesised into a scientific paper describing the same topics in a more coarse-grained and scientifically accurate way. The paper shall represent the main outcome of D2.2 and it will be submitted to the "PeerJ Computer Science" open access journal (classified as Q1 by Scimago up to 2020).

The HES-SO and UNIBO teams have designed and implemented a coherent framework able to integrate symbolic and sub-symbolic knowledge into a data-processing pipeline that combines knowledge-based (symbolic) and data-driven (sub-symbolic) predictors to produce recommendations and explanations based on generalized knowledge encoded in ontologies, raw data, and formal rules. The proposed pipeline extracts logic rules from data-driven predictors, employing the knowledge extraction methods DEXiRE and PSyKE developed by UNIBO and HES-SO teams and described in task T2.1 [wp2-1,wp2-2]. Then, the extracted rules are complemented with the knowledge encoded in ontologies and domain experts' rules to produce food recommendations with explanations. Finally, the pipeline re-injects rules in the sub-symbolic predictors to update and align their behavior to user and domain experts (e.g., nutritionists).

[T2.3 – M6-M10 (completed)] Semantic representation:

The application scenario characterizing the project is interdisciplinary and spans multiple concepts. Therefore, UNILU and LIST developed the first ontology-based negotiation for nutritional virtual coaches (NVC). The ontology is represented by OWL (Web Ontology Language). Protégé was used to represent the relationship between the users and their eating habits and to capture the structural similarities among the ingredients of the recipes. The ontology developed consisted of two main concepts: User and Food. The User concept captures the users' eating habits, such as religious or lifestyle restrictions. The Food concept involves a hierarchy of food recipe ingredients (e.g., Beef is a sub-concept of Animal Products, a Cucumber is a sub-concept of Vegetables etc.). This ontology has been later refined and extended by the HES-SO team.

To this end, the HES-SO team has focused on identifying existing ontologies to be extended and reconciled (linked) for the very use case (i.e., nutrition virtual coaching). The existing ontologies collection include food, user demographic, user behaviors, user preferences, nutrition goal/ambitions, persuasion theories &

techniques, health classes, ethical standing points, sustainability aspects, and legal aspects. The extension and integration of ontologies is taking place leveraging the software Protégé. The task was expecting to converge by M10. However, due to the difficulties in running the planned in-person focus groups involving external stakeholders (due to lockdown and related restrictive measure) expected in T4.1 and T5.3 this task has only recently been finalized. Such delay has minimally somehow impacted other tasks, whose convergency is under control, yet it might add some delays in the 3rd year. Moreover, the HES-SO team has designed and developed a zero-code tool running on a Web interface for nutrition professionals (i.e., doctors, nutritionists, and researchers) in the context of the underlying agent-based nutrition virtual coach. In particular, leveraging the underlying ontologies and persuasion schemas, professionals can define behavioral-change strategies per groups or single individuals. Such mechanisms will also enable the argumentation-based inter-agent explainability. Finally, HES-SO collaborated with UNIBO in designing and prototyping software extensions for PSyKE aimed at supporting interoperability among SKE and Semantic Web technologies (such as RDF, OWL, and SWRL). These extensions are described into the paper titled "Semantic Web-based Interoperability for Intelligent Agents with PsyKE" (see Section 2.1)

The final ontology illustrated in Figure 4.1 (see WP4, T4.1), represents and organizes the knowledge about the user characteristics, goals, food and exercise habits, lifestyle, preferences, ethical principles, medical conditions, food characteristics, nutrition virtual coach (NVC), and doctor agent. A complete description of the final ontology can be found in deliverables D3.1 and D4.1.

[T2.4 – M8-M14 (completed)] Intra-agent XAI library implementation:

The methods for SKE and SKI surveyed in T2.1 enabled the design and implementation of two (separate, yet interoperable) open-source software libraries for model-agnostic intra-agent XAI (cf. PsyKE and PsyKI). Other than enacting well-known principles for software engineering, separations allowed the UNIBO team to focus on SKI and SKE separately, hence working in parallel on both sides. Such a libraries are being integrated (3rd year) within the nutrition virtual e-coach platform (named EREBOTS).

To explain local (agent internal) DL predictors the UNIBO and HES-SO team have designed and implemented the above mentioned (T2.1) SKE-SKI libraries. Such libraries have been developed to be pluggable and usable by the most recent agent frameworks (i.e., SPADE). In particular, DEXiRE (to be integrated in EREBOTS in the 3rd year) executes the steps described in the Figure 2.1. It takes as input a trained DL predictor and the training set as inputs; extracts predictions from the predictor using the training data; the hidden neurons' activations are binarized; for each layer, the binary activation patterns are extracted; in each layer the most common activation pattern and the most decision-relevant neurons for each class are identified; intermediate rule sets inducing Boolean functions on the binary activation path are inducted; the intermediate rule sets are pruned to reduce the complexity of the final rule set; intermediate rule sets are expressed in terms of the input features; the final rule set is generated by merging the intermediate rule sets; and the final rule set is evaluated using the ground-true labels on the training set. The final rule set explains the internal decision process carried out by the DL predictor and can be transformed into an argument or natural language explanation [wp2-1].

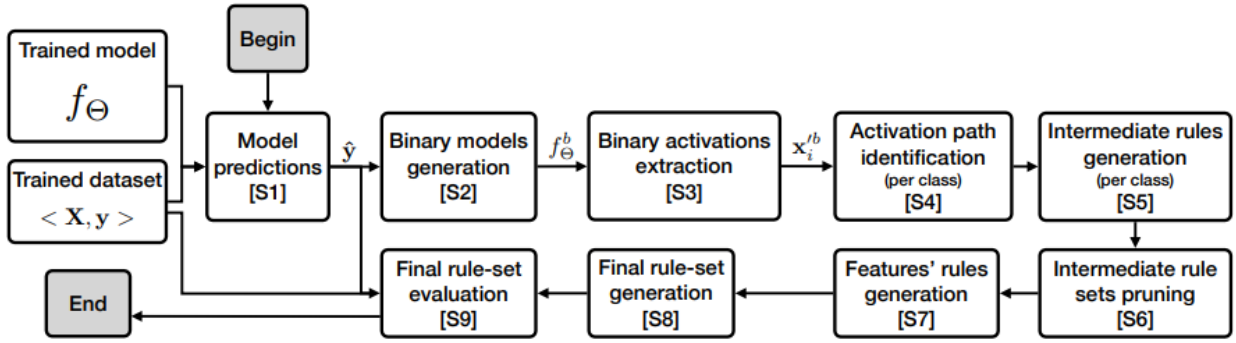

Figure 2.1 – DEXiRE algorithm step-by-step execution pipeline.

DEXiRE algorithm is described in the journal paper [wp2-t2.3-1] and has been applied to explain the behavior of a DL predictor in the context of head and neck cancer diagnosis [wp2-t2.3-2]. DEXiRE papers and software implementation are part of deliverables D2.1, D2.2, and D2.3.

[T2.5 – M10-M14 (completed)] Transparency assessment:

In line with the plan, the UNILU, OZU, and HES-SO teams have worked on transparency metrics. In particular, the three partners have investigated existing related metrics for transparency of explanations, and then have defined metrics to be used in our project. Such metrics will also be related and interconnected with the ontological schema characterizing the concepts within our system.

In particular, the teams have discussed the main concepts such as explainability, transparency and trust and review subjective and objective measurements of the effectiveness of explanations, based on the technology acceptance model (TAM). In general, the purpose of an explanation is to make the system's behavior **transparent:** the user's mental model about the purpose of the system, the system state, and how it works, should be in line with the actual system. Transparency is said to increase trust in the system, and improve usefulness and ease of use, and subsequently, the usage of the system. Specifically, we adapted the model of Hoffman et al [wp2-3] for the evaluation of explainable AI systems. We developed the model as shown in Figure 2.2. A more detailed discussion can be found in our publication [wp2-4].

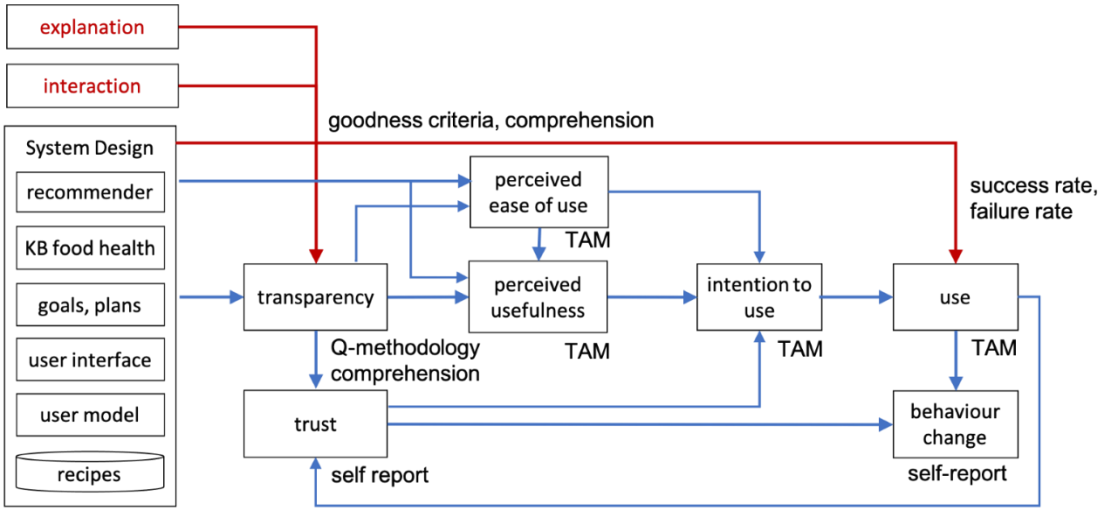

Figure 2.2 – Towards a Model for Evaluation of Explainable Recommender Systems [wp2-4]

The model consists of the following four components:

(a) **Goodness criteria** measure the success conditions, in this case of an explanation (see Table 1).

(b) **Test of satisfaction** tests "the degree to which users feel that they understand the AI system or process being explained to them." User satisfaction is measured by a series of Likert scales for key attributes of explanations: understandability, feeling of satisfaction, sufficiency of detail, completeness, usefulness, accuracy, and trustworthiness (see Table 1). Note that goodness criteria are meant for developers or researchers, whereas the satisfaction test is for end-users.

(c) **Test of comprehension** tests the effectiveness of an explanation on the user's mental model. This is similar to a kind of exam, so it may use open questions, or multiple-choice tests.

(d) **Test of performance** objectively tests over-all effectiveness of a system. In our case, that means that users indicate they like the recipe.

Table 1 below shows the goodness criteria for explanations, taken from Hoffman et al. [wp2-3]. Table 2 shows the survey questions we developed to evaluate the recommendations and explanations in our food domain [wp2-5], by analogy to the goodness criteria of Hoffman et al in Table 1.

1. The explanation helps me understand how the [software, algorithm, tool] works. [Y/N]
2. The explanation of how the [software, algorithm, tool] works is satisfying. [Y/N]
3. The explanation of the [software, algorithm, tool] sufficiently detailed. [Y/N]
4. The explanation of how the [software, algorithm, tool] works is sufficiently complete. [Y/N]
5. The explanation is actionable: it helps me know how to use the [software, algorithm, tool] [Y/N]
6. The explanation lets me know how accurate or reliable the [software, algorithm] is. [Y/N]
7. The explanation lets me know how trustworthy the [software, algorithm, tool] is. [Y/N]

Table 1. Goodness criteria for explanations [wp2-3]

1. The explanation/recommendation helps me to decide [which recipe to cook]. [Y/N]
2. The explanation/recommendation [which recipe to cook] is satisfying. [Y/N]
3. The explanation/recommendation [which recipe to cook] is sufficiently detailed. [Y/N]
4. The explanation/recommendation [which recipe to cook] is sufficiently complete. [Y/N]
5. The explanation/recommendation is actionable: it helps me to carry out my decision. [Y/N]
6. The explanation/recommendation lets me know how accurate or reliable the [recipe] is. [Y/N]
7. The explanation/recommendation lets me know how trustworthy the [recipe] is. [Y/N]

Table 2. Evaluation questions for explanation and recommendation dialogues in the food domain [wp2-5].

In our case, the recipe recommendation and explanation sessions are part of a long-term task, persuasion, to convince users to change their eating behavior and make it healthier. These sessions are part of the functionality of a Nutritional Virtual Coach (NVC). The purpose of an NVC is to induce behavioral change. To evaluate the effects of the persuasion task, we have to rely on repeated self-reporting experiments, see Figure 2.2. Based on earlier work on EREBOTS, the teams of HESSO and UNILU have are writing a paper on self-monitoring compliance to agreed plans, in order to evaluate the persuasion task.

References WP2

[wp2-1] Contreras, V., Marini, N., Fanda, L., Manzo, G., Mualla, Y., Calbimonte, J. P., ... & Calvaresi, D. (2022). A DEXiRE for extracting propositional rules from neural networks via binarization. Electronics, 11(24), 4171.

[wp2-2] Sabbatini, F., Ciatto, G., Calegari, R., & Omicini, A. (2021, September). On the Design of PSyKE: A Platform for Symbolic Knowledge Extraction. In WOA (pp. 29-48).

[wp2-3] Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for Explainable AI: Challenges and Prospects. arXiv:1812.04608 [cs.AI].

[wp2-4] Hulstijn, J., Tchappi, I, Najjar, A. & Aydoğan, R (2023). Metrics for evaluating explainable recommender systems. International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems (EXTRAAMAS 2023). D. Calvaresi (ed.), Springer Verlag: 212-230.

[wp2-5] Buzcu, B., Tessa, M., Tchappi, I., Najjar, A., Hulstijn, J., Calvaresi, D., Aydoğan, R. (to appear) Towards Interactive Explanation-based Nutrition Virtual Coaching Systems, 29 June 2023, PREPRINT (Version 1) available at Research Square.

chist-era

WP3

[T3.1 – M13-M20 (completed)] Symbolic data integration:

The HES-SO, UNIBO, and OZU teams have designed the interaction protocol between agents to exchange information in a heterogeneous environment. Agents can exchange messages containing knowledge (i.e., recommendations supported by logic rules). Such a knowledge can possibly be heterogeneous (requiring setting in place reconciling mechanisms such as [wp3-1]). To avoid such communication overhead, we have integrated and extended domain ontologies (i.e., food items, recipes, user model, and NVC). The exchange of punctual and aggregated information is supported by the implementation on the give NVC architecture (i.e., EREBOTS) of the reconciling data-schema. For more details, see D3.1 and D5.1.

[T3.2 – M17-M26 (ongoing)] Conflict resolution and multi-stakeholder reasoning on heterogenous data:

HES-SO, UNIBO, and OZU have designed and developed a bilateral negotiation protocol that enables recommendation agents to provide explanations regarding their recommendation and the user to give feedback to form a constructive dialogue between the user and the agent to resolve potential conflicts between the agents and the user. The protocol is depicted in Figure 3.1. The process begins with the user sending a recommendation request that includes their specific constraints (C), such as personal information and preferred cuisine. The recommender agent responds by providing a recipe (R) and an explanation (ε) that offers transparency into why the recipe was recommended. This allows the user to provide online feedback on the recipe, the recommendation, or both, giving the agent the chance to update the user's profile and improve future recommendations.

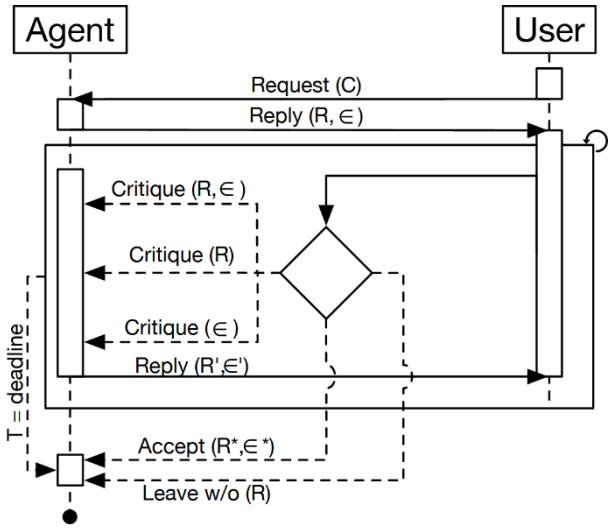

Figure 3.1 – FIPA description of the Negotiation Protocol.

Additionally, we have developed an agent to test this protocol. The agent retroactively generates explanations based on the recommendation criteria. For example, if a food item has a high protein count, the explanation generator selects a relevant sentence related to the protein percentage and adds it to a paragraph that contains a randomly chosen introductory sentence. This process is repeated for other recommendation criteria, resulting in a set of multiple explanations for each food item. The details of this agent and how the protocol was utilized can be found in our publication [wp3-2].

Additionally, we designed a more focused variant of the negotiation protocol above as illustrated in Figure 3.2. The key change in the protocol we are exploring is generating explanations per user request rather than sending them proactively. The design choices for further exploring this vein could be found in our EXTRAAMAS workshop publication [wp3-4].

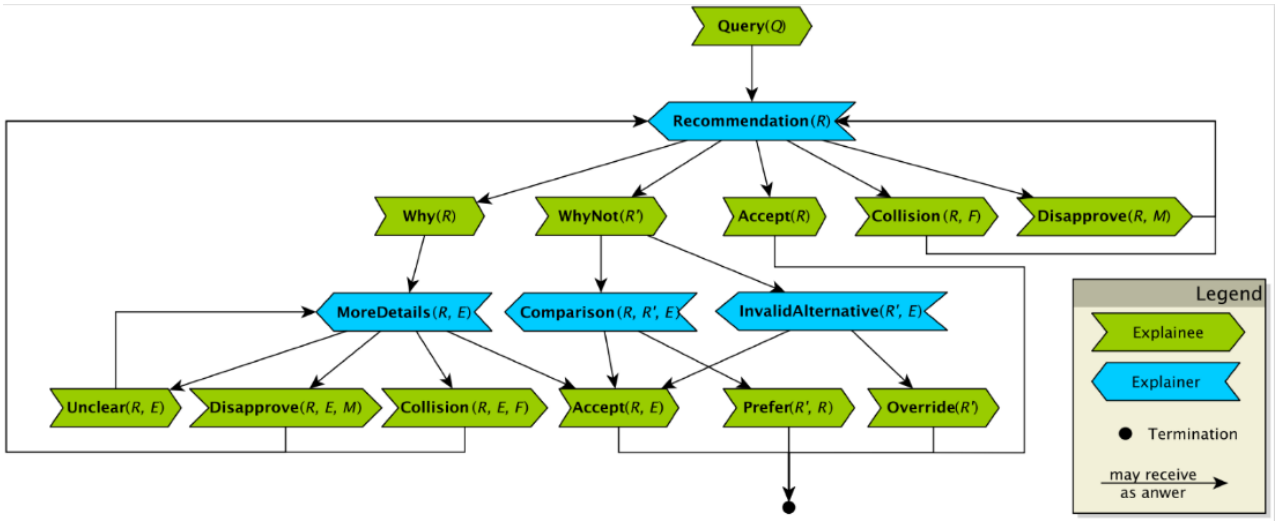

Figure 3.2 – Message Communication Diagram Between Explainer (Agent) and Explainee (User).

Furthermore, the partners involved in T3.2 built an explanation generation mechanism utilizing decision trees as they are often used by decision support systems given their simple and intuitive nature. They can explain the reasoning behind AI predictions or decisions in a more straightforward form than an otherwise black-box model. In order to discover the important features significantly influencing users' decisions (e.g., carbohydrates, protein, etc.), a decision tree is constructed from a labeled dataset. When we employ the user-based explanation generation method, the decision tree is constructed from historical data in which recipes are labeled with all users' decisions (i.e., accept or reject). Conversely, the item-based explanation generation approach utilizes the decision tree constructed from a set of recipes labeled according to the current user's constraints and feedback. For that tree, filtered and low-scoring recipes are negatively labeled (-1), recipes that align with the user's constraints are positively labeled (+1) and the rest is labeled neutrally (0). After sorting features with respect to their importance, we choose three of them to generate explanations for the given recipe. Additionally, we generate contrastive explanations, which are counterfactual explanations that compare the recommended recipe to an alternative that mainly highlights the recommendation, offering the system user transparency through comparisons. First, we select a recipe that is similar to the recommended recipe but its recipeScore is less than the recommended one. To do so, we utilize a pool of filtered (i.e., eliminated from the recommendation pool due to the user constraints/preferences) and/or low-scoring (i.e., not healthy or not tasty for the given user) recipes. We employ the Jaccard Similarity metric to determine the recipe similarity based on their ingredients. From this candidate set of recipes, we choose the one whose similarity with the current recommendation is maximum. Then, we compare features of the chosen counter recipe with those of the recommended recipe one by one. If the feature of the chosen recipe has a lower score for healthiness or user satisfaction, we added them into negative feature set.

From the features acquired by these methods, we generate a sentence using a predefined grammar-based structure to present them to the user. The structures are composed of two variants: one for the user / item-based explanations and the other one for contrastive explanations. The phrase repository of the system

consists of a set of phrases for each decision factor (e.g., for protein: "...provides sufficient protein..."), and other types of phrases such as subject and noun (e.g., "...this recipe..."). Finally, a detailed survey of existing explanation generation approaches can be found in our publication [wp3-3].

[T3.3 – M19-M26 (ongoing)] Evaluation of the heterogenous data integration framework:

In the NVC context, we envision a one-to-one (user-to-virtual agent) mapping. The user's personal agent (PA) employs knowledge-based and data-driven models using heterogeneous data sources (i.e., numerical structured information – calories, BMI, etc. and unstructured textual – recipes description, user' allergies, user's feedback, etc.) to produce recommendations and explanations. A common representation must be adopted to integrate these data sources consistently and semantically. In the case of the Expectation project, the partners have chosen to employ a symbolic representation based on logical rules (supported by ontologies) to provide explanations. This entails a reconciliation between the numerical and textual input data.

Therefore, recalling that the DL rules extraction tool (DEXiRE – T2.1 and T2.4) has been designed extracts rules from sub-symbolic models relating the input data to the expected output, in the case of unstructured data, it must cluster embeddings (numerical representation of unstructured data learnt by the network) to find patterns that can be expressed as a rule for that input. Finally, once the rule sets have been extracted, they can be efficiently integrated into the symbolic system.
To test and evaluate the above-mentioned approach, the HES-SO team is generating synthetic data that simulates a wide variety of users' inputs (numerical and behavioral) and contains conflictual information that must be processed and learned by data-driven predictors applying DEXiRE. To evaluate the quality of the assembled rules, the HES-SO team is designing an evaluation protocol that incorporates a set of performance metrics like accuracy, precision, recall, fidelity, and average execution time as quantitative evaluation. Additionally, in task T5.5, human feedback will be integrated as a quantitative and qualitative performance indicator.

[T3.4 – M24-M27 (ongoing)] Privacy and data management concerns:

The HES-SO, UNIBO, and UNILU executed a systematic literature review and focalized interviews to identify the legal and ethical concerns related to data privacy and security. As a result of this task, several ethical, societal, and legal challenges were identified, and for each one, the authors have proposed a set of mitigation measures to reduce the risk and the impact of those concerns; these results have been published in the journal paper [wp3-5]. With the challenges identified, the HES-SO team has extracted a series of software requirements that guided the architectural design and implementation of the chatbot platform EREBOTS in its current version 2.0. Among the architectural decisions taken to ensure users' data privacy and security, we remark on the use of the GDPR-compliant database Pryv to store users' personal data. Pryv database guarantees the users the ownership of their data and the ability to manage their access through a dynamic consent system. Additionally, access to the user's personal data has been limited to only their PA – the virtual coach. Although PA agents can share information and knowledge among each other, this is limited to compiled and aggregated knowledge and statistics, and rules extracted from the PA's predictors removing explicit references of their human users.

References WP3

[wp3-1] Stumme, G., & Maedche, A. (2001, August). FCA-Merge: Bottom-up merging of ontologies. In IJCAI (Vol. 1, pp. 225-230).

[wp3-2] Buzcu, B., Varadhajaran, V., Tchappi, I., Najjar, A., Calvaresi, D., Aydoğan, R. (2023). Explanation-Based Negotiation Protocol for Nutrition Virtual Coaching. PRIMA 2022. Lecture Notes in Computer Science, vol 13753. Springer, Cham. https://doi.org/10.1007/978-3-031-21203-1_2
[wp3-3] Berk Buzcu, Melissa Tessa, Igor Tchappi et al. Towards Interactive Explanation-based Nutrition Virtual Coaching Systems, 29 June 2023, PREPRINT (Version 1) available at Research Square
[wp3-4] Ciatto, G., Magnini, M., Buzcu, B., Aydoğan, R., Omicini, A. (2023). A General-Purpose Protocol for Multi-agent Based Explanations. Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2023. Lecture Notes in Computer Science vol 14127. Springer, Cham.
[wp3-5] Calvaresi, D., Carli, R., Piguet, J. G., Contreras, V. H., Luzzani, G., Najjar, A., ... & Schumacher, M. (2022). Ethical and legal considerations for nutrition virtual coaches. AI and Ethics, 1-28.

WP4

[T4.1 – M5-M10 (completed)] User modelling:

Given the nutrition-related context, the HES-SO team has collaborated with a professional nutritionist in identifying the stakeholder classes (i.e., people with chronic diseases, food disorders, weight issues, food-related autoimmune diseases, gastric problems, wellbeing, sport, sustainable lifestyle oriented, nutrition experts, food explorers, on budget, families with children) and characterization, and their possible correlation with the use of nutritional virtual coaches. Moreover, to have a more comprehensive understanding of the users, we have realized two questionnaires (~60 and ~10 questions) that we are running as online forms. The investigation protocol foresees also several in-person meetings including stakeholders and expert nutritionists. Unfortunately, due to COVID-19-related restrictions, we had accumulated some delay to organize them (yet, as of today, the consortium has completed the task). Finally, given the background and interests of the OZU team in creating models for predicting user's interests/preferences via machine learning, they took part into this task. In particular, HES-SO and OZU have collaborated in identifying the key features of the user's interests/preferences meaningful to generate recommendation and arguments (explanations). OZU evaluated the applicability and performance of the proposed preference elicitation approach. To evaluate to what extent the designed modeling approach is practical in terms of accuracy and cognitive load on the users, OZU, HES-SO and UNILU designed and conducted user experiments where participants can interact with the agent through a user interface and give feedback on the displayed food recipes in order to evaluate to what extent the designed modeling approach is practical in terms of accuracy and cognitive load on the users.

OZU first built a Turkish food recipe dataset consisting of various recipes and ingredients. Each recipe had the following features: category, ingredients, cuisine, cooking time, calorie, carbohydrate, fat, protein, fiber values. While interacting with users, the system displays categorical values such as `Low', `Normal', and `High' instead of exact values of nutrition to help users specify their preferences about nutrition levels.

To study whether the system could predict the users' preferences on the food recipes (i.e., either like or dislike) based on the feedback on very few recipes, we conducted a user experiment. Our experimental setting consisted of four phases as follows:

(a) Binary Preference Elicitation on Diverse Recipes
(b) Explanative Preference Elicitation on Diverse Recipes
(c) Binary Evaluation of the Learned Preference Model
(d) Likert-Scale Evaluation of the Learned Preference Model

First, the OZU team applied the K-Medoids Clustering technique for achieving a diversity sampling strategy; consequently, randomly chosen recipes from each cluster are displayed to the user. As shown in Figure 4.1, they were asked to indicate whether they like or dislike the displayed food recipes. Based on their feedback,

Logistic regression and Graph-based Sampling are used as the Active Learning strategy to predict their food preferences (i.e., whether the given food recipe is liked or disliked by the user).

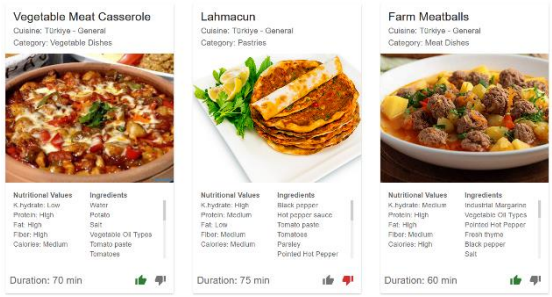

Figure 4.1 – Binary Preference Elicitation Interface

In the second phase, users are asked to specify structured explanations for their preferences on a chosen recipe by indicating the factors (i.e., ingredients, cooking, cuisine) affecting their choices positively and negatively as well as the factors not having any influences on their choice as visible in Figure 4.2. Since this process may increase the participants' cognitive load significantly, only five recipes are examined in total. Accordingly, the learned model is updated based on the users' inputs.

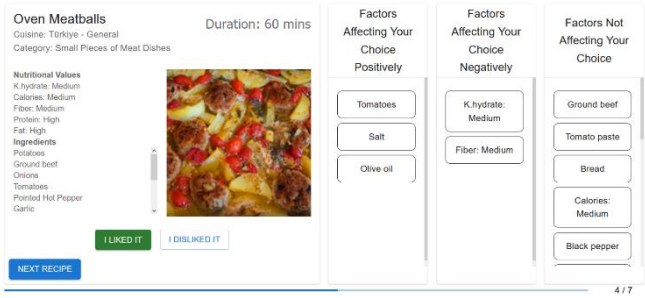

Figure 4.2 – Explanative Preference Elicitation Interface

The third and fourth phases are the evaluation of the learned preference model. In the third phase, we ask users to indicate their binary preferences (i.e., whether they like or dislike) on 25 recipes in a similar interface, as seen in Figure 4.2. Five recipes are chosen from the food recipes shown in the first phase to test the users' consistency. While analyzing those results, if we detect any inconsistency in the users' responses, we do not consider those users' data in our final evaluation. The remaining 20 recipes are selected using a testing sampling approach where we take the instances where the model predictions are the most certain and discard similar instances. Note that the similarity between food recipes are measured with the Jaccard similarity. Note that the average similarity within a cluster was around 0.2; therefore, we consider if recipes are similar if their Jaccard similarity is above 0.25. In this way, we can ensure the diversity of test samples which provides a more general scoring over the food dataset than only scoring top predictions for the participant which can include very similar recipes. Similarly, we pick ten food recipes and denote whether the system labels the user will ``like'' or ``dislike'' in the fourth phase. The users specify to what extent they agree with the system decision on a 5-point Likert scale (i.e., strongly/moderately disagree, neutral, moderate/strongly agree).

For human experiments, the required user interfaces and algorithms have been implemented. Before starting the experiments, we explained the experiment procedure to all participants elaborately and got their consent to participate in our experiments. In addition, they watched a short video demonstrating how they should interact with the system before each phase.

The OZU team ran a study that evaluated the effectiveness of various AL strategies and the trade-off between the users' human effort and the quality of the elicited user profiles. Our results showed that our strategies

outperformed the baseline, particularly when user feedback is further integrated into the preference models, showcasing the effectiveness of the system in the long-term. You can find a more detail analysis of our result in our paper [wp4-1]. Our findings also showed that the users found the task of manually labeling ingredients as a preference explanation form to be exhausting. However, they rated these systems as highly explainable. This would suggest that there is a trade-off between system effectiveness and human-effort that needs to be optimized.

The HES-SO and OZU teams have identified the most relevant user features required to provide personalized food recommendations and explanations. These features constitute the user static model and have been integrated into the final ontology. Figure 4.3 illustrates a graphical representation of the final ontology, where the user entity represents the user model in the system. The user model is composed of the primary user profile that identifies the user and the user dimensions that provide the user context. In the primary user profile, we can find the essential features required to produce accurate nutrition plans and food recommendations, like age range, medical gender, lifestyle (activity level), marital status, weight, and height. In the user dimensions, we have identified eight fundamental user aspects that define the user context:

- **Cultural factors:** this dimension is essential to identify users' food restrictions based on cultural (i.e., vegetarian, vegan) and religious (kosher, halal) factors.
- **Sustainability:** The sustainability dimension describes the user's concerns about the environmental effects of the food recommendations
- **Ethical principles:** Ethical principles characterize the user's ethical concerns and establish an ethical framework to evaluate the food recommendations, find ethical violations, and learn from the user's feedback to align recommendations with the user's values and moral code.
- **Actions:** Those are the actions that users can execute on the system. Actions briefly describe the available interactions between users and the system.
- **Preferences:** This dimension models the user's choices related to mealtime and is essential to synchronize recommendations.
- **User goals:** This dimension describes the user's goals, which the NVC employs to produce plans and actions to achieve those goals.
- **Health conditions:** Health conditions provide a holistic view of the user's physical and mental state. The *Health Conditions* entity is essential to plan treatments and give or avoid certain nutrition items.
- **User category:** The user category dimension is defined by the doctor or domain expert (i.e., nutritionist), and it is employed to define the appropriate treatment and filter food recommendations according to the user's prototype.

A detailed description of the user model and the final ontology can be found in the deliverables D3.1 and D4.1.

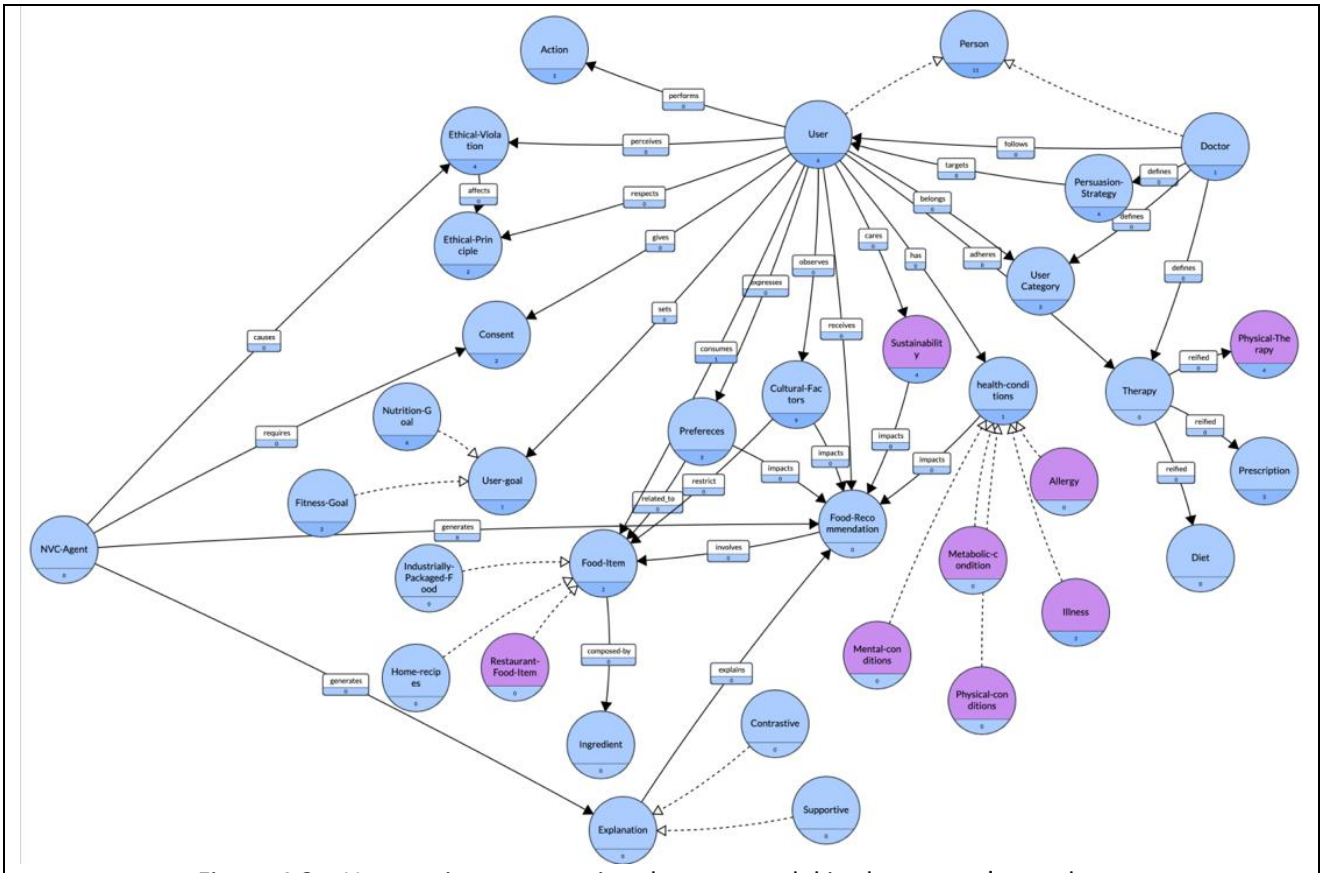

Figure 4.3 – User entity representing the user model in the system's ontology.
(for a detailed on the ontology description see deliverable D4.1)

[T4.2 – M8-M11 (completed)] Interaction Protocol:

The OZU team has collaborated with the UNILU and HES-SO teams to design 1-to-(1-to-n) negotiation protocol governing the user-agent and agent-agent(s) interactions (prototyping a dedicated framework).

To do so, the teams have co-designed the interaction protocol enabling inter-agent explainability and particularly focused on human-agent interaction. This interaction protocol brings out the communication between one given user and their personal food coaching agent as well as the explanation and motivation associated to the recommendation. Moreover, they have identified open issues of the interaction protocol based on intra-agent explainability. In this scenario, two main situations are considered: centralized interaction protocol and decentralized interaction protocol. Each of them having advantages and disadvantages. For example, centralized approach has the advantage to be easy to model; however, the central point is critic. The decentralized approach has the advantage to scale. However, it increases the negotiation's complexity.

Other activities carried out in this task include a survey of recommendation systems in the food domain, and a survey of Argumentation-based Negotiation approaches. As of today, the survey has been redacted in the form of technical reports. We envision to write a wrap-up paper when the including the results of T4.1.

The HES-SO team has designed the interaction protocol between the user and their PA, allowing the user to execute the following actions with the system:

- Ask for a recommendation on food or restaurant recommendation.
- Ask for an explanation of the given recommendation.
- Provide feedback on the recommendation, explanation, or on the whole application.
- Create an account and a profile.
- Delete the account and the profile.
- Provide or revoke consent to access their private data.
- Track their progress and food consumption employing multimodal data (text, images, and audio).
- Check summarized statistics.

- Receive proactive recommendations and notifications to remind them to follow healthy eating habits.
- Negotiate with the PA for a plan or a recommendation.

[T4.3 – M10-M20 (ongoing)] Agent-based profiling:

The OZU team has surveyed existing explainable recommendation and preference elicitation techniques. The OZU team first focused on learning users' preferences on food attributes (e.g., whether the user (dis)likes given ingredients/combinations) and developed an active learning approach for eliciting such preferences. While facilitating the development of this preference elicitation technique, they manually created some synthetic users to assess the effectiveness of the underlying approach and consequently to revise the approach accordingly. The UNILU team has started to model the user behavior. The user is characterized by its age, sex, ethnicity, etc. (full profiling obtained by T2.3 and T4.2). Moreover, they started to model the agent behavior. The agent is rational that seeking to maximize the utility function governing its behavior. This work is still ongoing.

An important component in conflict resolution in recommender systems using negotiation is preference elicitation. Therefore, we have developed and tested various Active Learning based strategies to model the user's food preferences. This model was used to explain as to why a user would prefer or not prefer a given recipe, and was detailed in WP4, T4.1.

The agent-based profiling is a dynamic data-driven process executed as the user interacts with the chatbot platform. In each interaction, the user can provide explicit feedback about the recommendations, explanations, plans, arguments, or the platform itself. Explicit feedback is employed to update the rules, knowledge bases, and machine learning predictors, achieving a deeper personalization level by adapting the PA to the user as the interaction evolves. Additionally, agent-based profiling uses the tracking information from the user to profile their behavior and their progression toward their goals. Agent-based profiling characterizes the dynamic user's food preferences, meal consumption habits, and activity patterns that are employed to provide better recommendations and explanations.

[T4.4 – M15-M22 (ongoing)] Multi-modal explanation communication:

UNILU and LIST developed a multistep approach for explanation generation. The first step is the feature importance using SHAP (both feature importance of recipes is used). The second step is the generation (both plain explanation and contrastive explanation can be generated). UNILU and LIST developed two main methods for the explanation generation:

(a) a predefined method using a predefined grammar. A plain and contrastive grammar is defined in order to enrich the user experience when interacting with the NVC.
(b) a more flexible approach using prompt engineering and LLMs (e.g., GPT, Flan T5) to generate the templates in (a) – to be further revised.

Furthermore, UNILU and LIST developed (currently developing) the interfaces for the interaction between the users and the system. The first interface developed (currently developing) was using QT robot and the second is using Furhat robot.

The HES-SO team has developed the rule extraction tool DEXiRE (described in T2.4), which produces a set of rules as the explanation of the DL predictor decision process. Rule sets allow further transformation to produce explanations in natural language that could be presented to the user as a short textual explanation or synthesized as an audio message that can be reproduced by a robot or sent in an instant message application. Additionally, DEXiRE can produce logic diagrams describing the activation path for each output

class, which can complement the textual and audio explanations, generating multimodal, complementary descriptions of the decision process carried out by the virtual nutrition coach.

[T4.5 – M19-M31 (ongoing)] Ethical concerns about data privacy:

As part of the effort to identify and analyze ethical and societal concerns (Task 5.2), UNILU, HES-SO, and OZU teams have also discussed data privacy issues, see publication [6]. Generally, food is a sensitive issue, also in the sense of the special data categories of GDPR article 9, because food relates to personal or cultural identity, religious practices (kosher; halal), or membership of ethnic minorities. In addition, food and especially diets relate to illness and health, which are also a special category.

When we collect data about user preferences or when we use data sets for training preference models, we must take ethical concerns about data privacy into account. One can look at these concerns from two perspectives: (1) legal, focusing on the protection of personal data, under the GDPR, and (2) ethical, focusing on the responsibility we have as system designers towards our users and subjects.

Regarding (1), all our user studies take place under supervision of the ethical committee of the university involved, that judges the acceptability of the research design of the studies relative to their purpose. As part of the study protocol, users are asked to give informed consent, for the use of their personal data for scientific purposes (GDPR article 9.2(a). All data related to user studies are stored in a secure environment. As a rule, data sets are not shared between the universities, unless they are anonymized. In addition, where we use publicly available data as training sets, these data sets are first anonymized and afterwards tested not to contain ana remaining personal data.

Regarding (2), the entire research project is about developing explainable recommender systems. One of the purposes of explainable AI is to improve transparency of the system. That means that users and subjects develop a mental model of how the systems works and for what purpose, that is aligned with reality. In general, transparency allows users to critically examine the system and determine to what extend it can be trusted. So that means that explanations are part of the solution.

Specifically, the system must be able to explain to the user or subject why it needs to collect sensitive user preferences, namely, to provide a better recommendation. Here we have a trade-off: if we collect more sensitive data, we can potentially provide a more fitting recommendation, but that also increases the risk of a breach of confidence. To test if we get the balance right, we can add an evaluation metric (task 2.5), that calculates the amount of sensitive data collected, relative to the complexity of the recommendation given. This measures if sensitive data are necessary for the recommendation.

References WP4

[wp4-1] Cantürk, F., & Aydoğan, R. (2023). Explainable Active Learning for Preference Elicitation. *arXiv preprint arXiv:2309.00356*.

WP5:

[T5.1 – M1-M5 (completed)] Architecture requirements:

The HES-SO and UNIBO teams collaborated into refining the requirements and sketching the architecture of the proof-of-concept system targeted by WP5. Moreover, both the teams collaborated into sketching the conceptual and software architecture of the prototype (whose actual development is currently ongoing). Moreover, the teams are about to conduct in-person focus groups (delay due to covid-19-related restrictions) to involve them into the cocreation of some features/interfaces and elaborate the expectations and related metrics.

- Analyzing the information obtained from surveys and focus-group meetings, the HES-SO and UNILU teams have elicited the architectural requirements of the proof-of-concept system. The requirements have been divided into categories according to the needs they satisfy. The categories identified are as follows:
- **Software requirements** specify the architectural and technological choices and the quality attributes of the proof-of-concept system. The software requirements are further divided into the Functional (FR) and Non-functional (NFR) requirements subcategories, defined as follows:
    - **Functional requirements (FR):** describe the system's expected behavior and functionality. The FR identified for the system to be released in this project concern front end, proactive behaviors, multi-modal communication, user-agent exclusive mapping, front-end – backend decoupling and modularity, thematic behaviors, persuasive strategies, agent-agent communication, personalized food recommendations, personalized restaurants recommendations, food tracking, data synchronization, user registration, user login/logout, dynamic consent management, reception of general feedback.
    - **Non-functional requirements (NFR):** These requirements go beyond the system's basic functionality and focus on the quality attributes necessary to meet stakeholder expectations. In particular, the HES-SO team has identified data integrity, multi-agent-based architecture, data retention, self-contained deployment, disaster recovery, extensibility, flexibility, integrability, internationalization and localization, interoperability, front-end compatibility privacy, portability, security, transparency,
- **Ethical requirements (ER):** Respond to ethical challenges identified in task T4.5. (For more details, see task T5.2 and deliverables D5.1).
- **Societal requirements (SR):** Respond to the societal concerns raised from focus group interviews. (For more details, see task T5.2 and deliverables D5.1).

See D5.1 for more details on FR, NFR, ER, and SR.

[T5.2 – M3-M12 (completed)] Identify ethical and societal concerns:

The HES-SO, UNILU, and LIST teams have defined the proposed system as a nutrition virtual coach (NVC). The scope of an NVC goes beyond traditional recommender systems, since it seats at the overlap of informative systems, persuasion techniques/systems, argumentation techniques/systems, and personalized assistive systems. Therefore, the NVCs' ethical and legal challenges must take those four dimensions into consideration. The two teams have elaborated on both domain-specific and cross-domain challenges/criticalities and have identified possible strategies of mitigation and countermeasures to the elicited pitfalls of conventional systems. Such elaborations are redacted in a form of technical report, and it is currently being shaped into a journal paper. The journal targeted is still under discussion among the participating teams. Nevertheless, the priority is to have it open access and Q1/Q2.

HES-SO, UNILU, LIST, and OZU have developed a questionnaire to gather opinions from people of all ages, genders, professions, occupations, races, ethnic groups, and cultures, including experts in the field of nutrition about NVC. To this end, an online and anonymized form was used. The online form was divided into 5 main sections, named personal experience of the virtual nutritional coach system (to describe the past experiences with a nutrition coach of participants), the concept of an explanation (the content of a good explanation), Confidentiality (protection of participants sensitive data), trust and personal information of the user. About 140 persons filled out the questionnaire.

Moreover, as Nutritional Virtual Coaches (NVC) like mostly all tech-based solutions come with their own set of societal implications and responsibilities, HES-SO, UNILU, LIST, and OZU have identified some societal requirements for the NVC application. In fact, NVC are designed to suggest dietary choices to users, and therefore they are not merely about algorithms and databases. They hold a mirror to the diverse, intricate societal fabric where they operate. To ensure that our proposed system are universally beneficial and ethically grounded, certain social requirements must be prioritized.

One of the first social requirements is user privacy and data security. Users entrust NVC platforms with sensitive information, expecting discretion and protection. Alongside, it's crucial to emphasize inclusivity, ensuring the system caters to the dietary needs, cultural backgrounds, and health conditions of its audience. Furthermore, the integrity of recommendations, grounded in evidence-based research, cannot be compromised. This is closely tied to the system's responsibility to combat misinformation, which is rife in the realm of nutrition.

In a modern world, cultural sensitivity becomes paramount, recognizing and valuing diverse food traditions and norms. Moreover, the balance of power must tilt in favor of user autonomy, allowing individuals the freedom to make informed decisions without undue coercion. As we navigate the intersection of technology and nutrition, these social requirements not only enhance the utility of food recommender systems but also uphold the very essence of ethical digital innovation.

Furthermore, designing such a system, involving diverse users in the design and testing phase ensures the platform's inclusivity and effectiveness. Feedback loops, constant iteration, and user-centric design thinking are crucial. In fact, user studies are instrumental in understanding the societal requirements of a system, ensuring that it aligns with the needs and values of the broader community.

Overall, the SR present identified are ensure privacy and security, accessibility, social interaction features cultural sensitivity, feedback and adaptability, diversification in needs and preferences, educational components, emergency protocols, socioeconomics considerations.

The HES-SO and UNILU teams have finalized the task to identify ethical and societal concerns, and completed the report, which was successfully published (see [wp5-1]).
These ethical and societal issues have also been considered in the development of the metrics for evaluation (see Task 2.5). Specifically, some ethical concerns (e.g., data quality, manipulation) can be reduced by interactive explanations, which improves transparency of the system, and allows users to critically assess trustworthiness of the recommendations.

Since Nutrition Virtual Coaching (NVC) systems could directly impact the psychological and physical health of the users, several ethical concerns have been raised, identified, categorized, and expressed as ethical challenges (EC). The ethical requirements were defined in relationship to ethical challenges (EC); in the Expectation project, the ethical challenges were distributed among the NVC main submodules as follows:

Food recommender system: health and moral damaging, privacy, personal identity threat, opacity, biased recommendation, social pressure.
Argumentative systems: attain formal validity, leverage sole sincerity/truth, ensure content justice, enact fair and just procedures, compliance-verification coverage, simplify or aggregate arguments, multimodal arguments.
Informative and assistive systems: facilitate technology access, ensure the system identity, ensure medical data confidentiality, affordability of the proposed solutions, to ensure safety boundaries.
Persuasive systems: persuasion/nudges awareness, clear goals statements, prevent unintended behavior change.

Based on ethical challenges identified before, the following ethical requirements (ER) have been defined: appropriate recommendations, privacy, identity protections, transparency, bias mitigation, social pressure mitigation, argument validity, fair argumentative procedure, argument coverage, simple arguments,

multimodal communications technology access, personal agent identity, system boundaries, clear goals, preventing unintended behavioral change.

[T5.3 – M5-M16 (ongoing)] Data collection:

The OZU team has investigated existing food ontologies and datasets. They have also examined some datasets to procure an initial prototype and determine which data properties are vital. A list of the datasets we have examined follows: FoodKG - A knowledge graph that contains the recipe, ingredients, cooking steps, and nutritional values; FoodOn - An OWL ontology with various food properties; BBC Food Ontology; Recipe1M+; WikiProjectFood/Taxonomy. The following websites can also be utilized to gather recipe related information: https://www.allrecipes.com/; http://www.foodsubs.com/; https://www.yummly.com/; https://www.food.com/

UNILU and LIST proposed to use the dataset foodRecSys-V1 from the well kwon website http://allrecipes.com available in Kaggle platform. This dataset was used in one of the experiment done by OZU and the finding disseminated in a paper. The dataset is composed of 52,821 recipes from 27 categories posted between 2000 and 2018. After preprocessing, the dataset is composed of rating of 1,160,267 users, 49,698 recipes with 38,131 ingredients and 3,794,003 interactions.

Together with HES-SO and UniBo, the OZU team started working on a survey preparation about argumentation-based negotiation approaches. Moreover, the HES-SO team is working on the platform designed in T5.1 to allow the inclusion of the several sub-components and the "virtual" knowledge encoded by this task. Soon a first deployment of the system EREBOTS 2.0 will take place, thus allowing the HES-SO team to collect real-user data and enable the autonomous agents' active learning designed by the OZU team.

As part of the effort to develop metrics for evaluation (Task 2.5), UNILU, LIST, UNIBO, and OZU teams have discussed and decided on the main conditions to be tested in the user experiments of WP5. This has resulted in a general research model (see Figure 1, under task 2.5). Apart from the usual testing of functionalities, three conditions are evaluated: system without explanations, system with one-shot explanation, and system with interactive explanations. Initial results of these developments have been submitted to a journal for publication [paper-to-appear].

Before that, the OZU team has further developed and tested the GUI version of the recommender system [paper-to-appear]. UNILU is now busy developing a 'virtual human' interface, which shows a realistic looking talking head. So at least three versions of the system can be compared, to test hypotheses about the effect of the embodied nature of a social robot system: (1) GUI, (2) virtual human and (3) social robot.

The HES-SO team has been enriching the recipe database with queries to different information sources available on the internet, employing targeted questions directed to satisfy the needs of several cultural groups (e.g., vegetarian, kosher, halal, etc.). Additionally, the recipes have been categorized according to possible allergenic factors and the number of calories per portion to facilitate their integration with the ontology and their use within machine learning models to generate recommendations for different target groups.

[T5.3 – M5-M16 (ongoing)] Architecture Implementation:

Based on the requirements identified in task T5.1 and described in deliverable D5.1, the HES-SO team has designed a multi-agent chatbot platform named EREBOTS, currently under development in its second version. EREBOTS v2.0 is designed to be front-end agnostic and platform-independent, ensuring high flexibility and integrability with different front ends and instant messaging applications such as Telegram, WhatsApp, and Messenger. A detailed architectural description of the EREBOTS 2.0 system can be found in deliverable D5.1.

Figure 4 illustrates the technological and deployment point of view of the EREBOTS 2.0 system. As is shown in the figure, the different components of the system are isolated in docker containers that communicate

with each other using the internal network provided by docker. The multi-agent system (MAS) is built on a SPADE multi-agent platform and employs the XMPP server Prosody to exchange messages between agents. The private data is stored on the GDPR-compliant database Pryv, and the configuration, statistical, and public data is stored in a MongoDB database hosted in its own container. Front-end custom app HemerApp (custom developed front-end) and doctor agent front ends are hosted in separate containers or distributed as mobile or desktop applications.

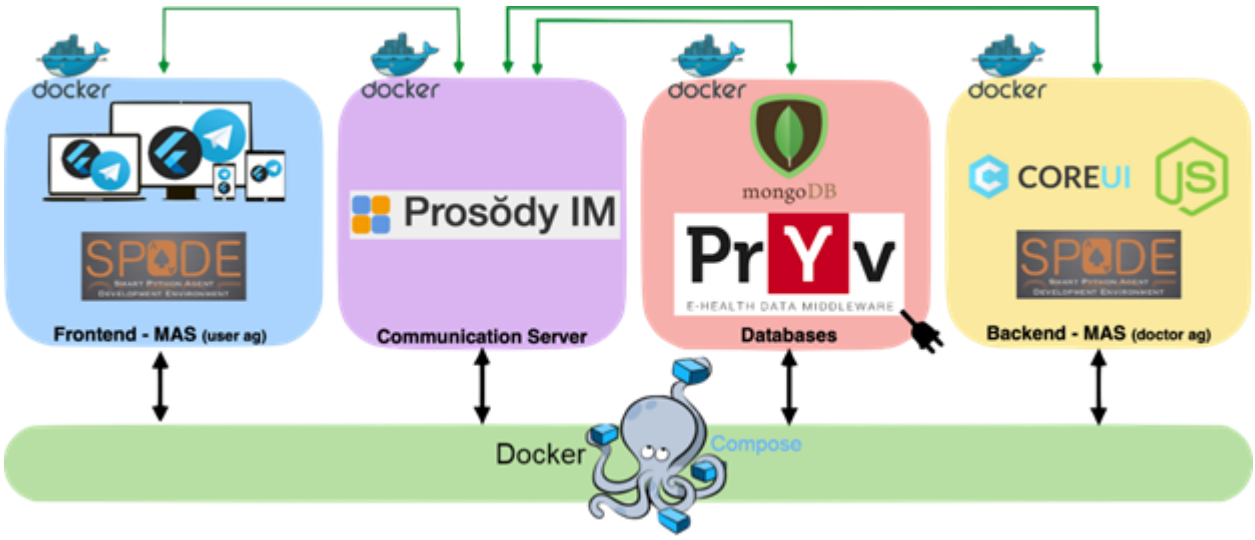

Figure 5.1 – Deployment and technology diagram of EREBOTS 2.0. The containers composing the EREBOTS system from left to right: Frontend container, Prosody XMPP server, database server, and Spade multi-agent system backend.

To facilitate the information exchange between agents, the HES-SO team is implementing the multi-agent chatbot platform EREBOTS 2.0, which supports the message exchange between heterogeneous agents. As is illustrated in Figure 5.2, each user has assigned a patient agent (PA), which is a personal virtual entity that has access to the specific user's data to provide that particular user plans, recommendations, and explanations behaving like a personal nutrition coach for that user. Additionally, the PA agent implements three behaviors: thematic, persuasive, and routing message. The thematical behavior produces food recommendations, plans, and performs the user's tracking. The routing behavior manages the incoming messages and addresses them to the compelling behavior. Persuasive behavior implements the strategies required to encourage the user to change their habits to improve their health. The PA agent can communicate with the user and with other PA agents through a standardized message format.

To make the EREBOTS platform independent of the front-end or client applications, all the external communication is mediated by the Gateway Agent (GA), which adapts the messages' formats to/from the given front end and the targeted instance of the platform.

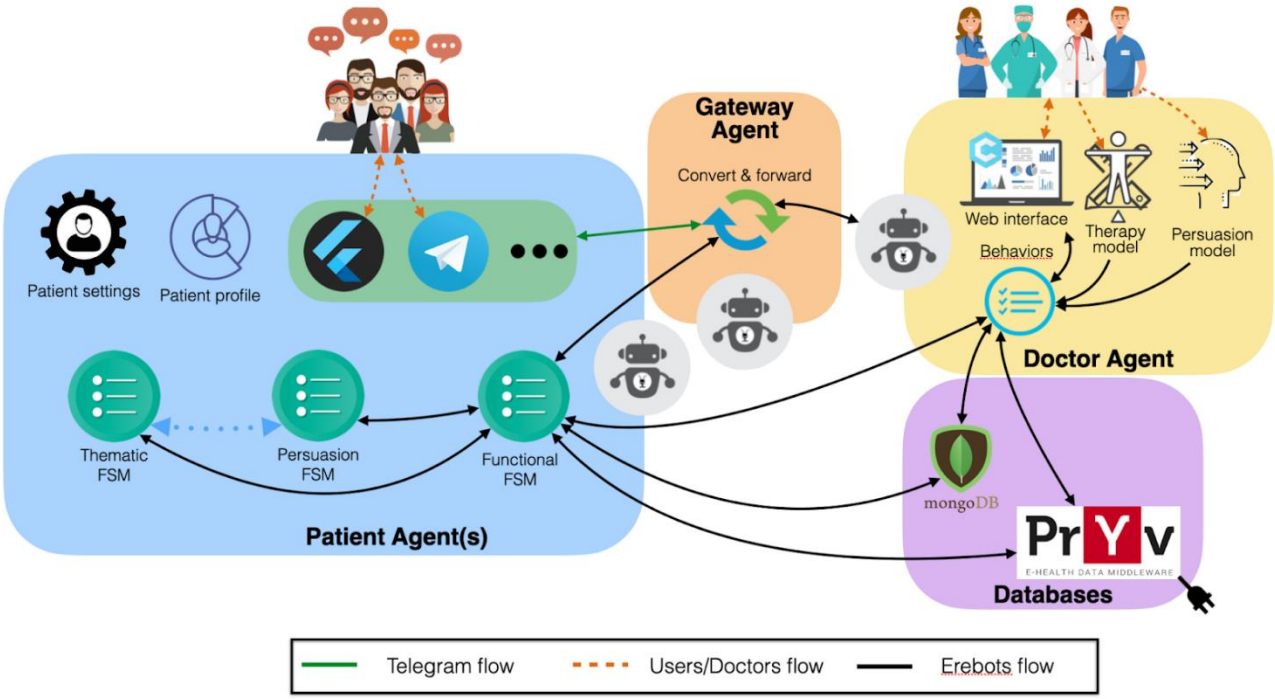

Figure 5.2 – EREBOTS v2 architectural platform design.

Figure 5.2 illustrates the EREBOTS v2 logic component and their interaction. From left to right, the profile screen, the chat screen executing the profiling task on the HemerApp front-end, and the food recommendation interaction between the user and the PA agent.

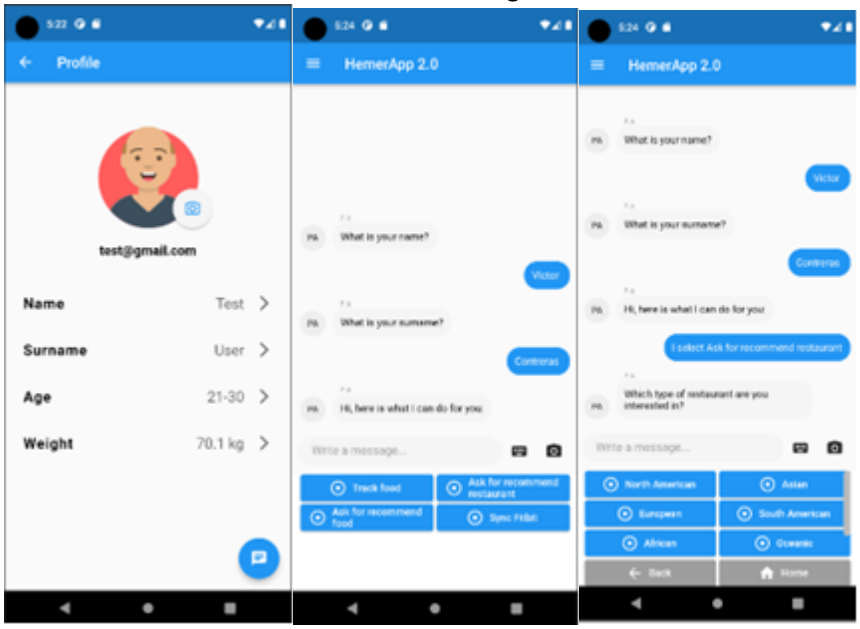

Figure 5.3 – EREBOTS's front end (named HemerApp): from left to right: i) User profile ii) Chat main screen iii) chat thematic screen.

Alongside of the EREBOTS platform, the OZU team has developed a lighter system to test their recommender engine and negotiation strategies (see T3.2) for providing valuable explanations.
Leveraging the OWL ontology (seeT2.3) and the explainable negotiation protocol (see T3.2), a human experiment consistent of two sessions where a simple recommendation strategy was compared to our approaches was conducted via an online system with 53 participants from Turkey. The results have been publicized in the PRIMA conference. Figure 5.4 shows the user interface developed to accommodate the interactive aspects of the protocol, where the user can give feedback and receive explanations with recipe related information.

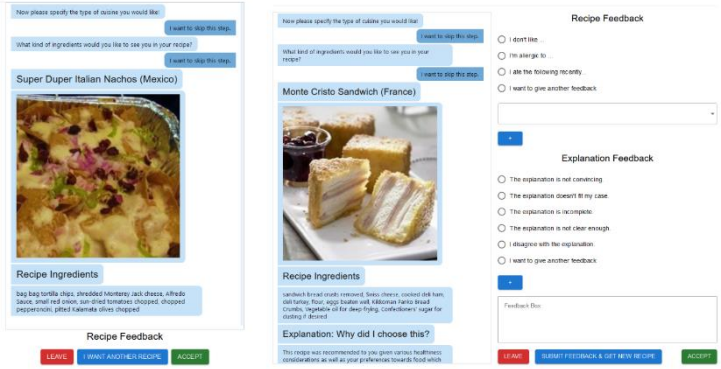

Figure 5.4 – Regular and Interactive Recommendation User Interfaces.

Following the study, we acquired some feedback for experiment attendees. In light of the feedback received and our latest explanation generation techniques base (The user/item based explanations and contrastive explanations, see D3.2). Summarily, the user/item-based explanations are alluring sentences about each positive feature. They are intended to be brief and pithy, whereas contrastive explanations aim to create a comparative explanation with a worse alternative (which can be longer). To realize the extension of explanation generation mechanisms for a human experiment setting, we had to update our food recommendation interface. We improved how the food recipes, and their supportive explanations are displayed to communicate the explanations more effectively and diminish the effect of factors irrelevant to the quality of explanations, such as pictures. Nutritional information and main ingredients are shown directly alongside several types of explanations. Additionally, we have also updated the presentation of the explanations. User Experience is an important aspect for the explanations given that users are often “bored” out of reading them, per the feedback we received in the earlier experiments. Toward this end, Figure 5.5 shows the new interactive recommender system.

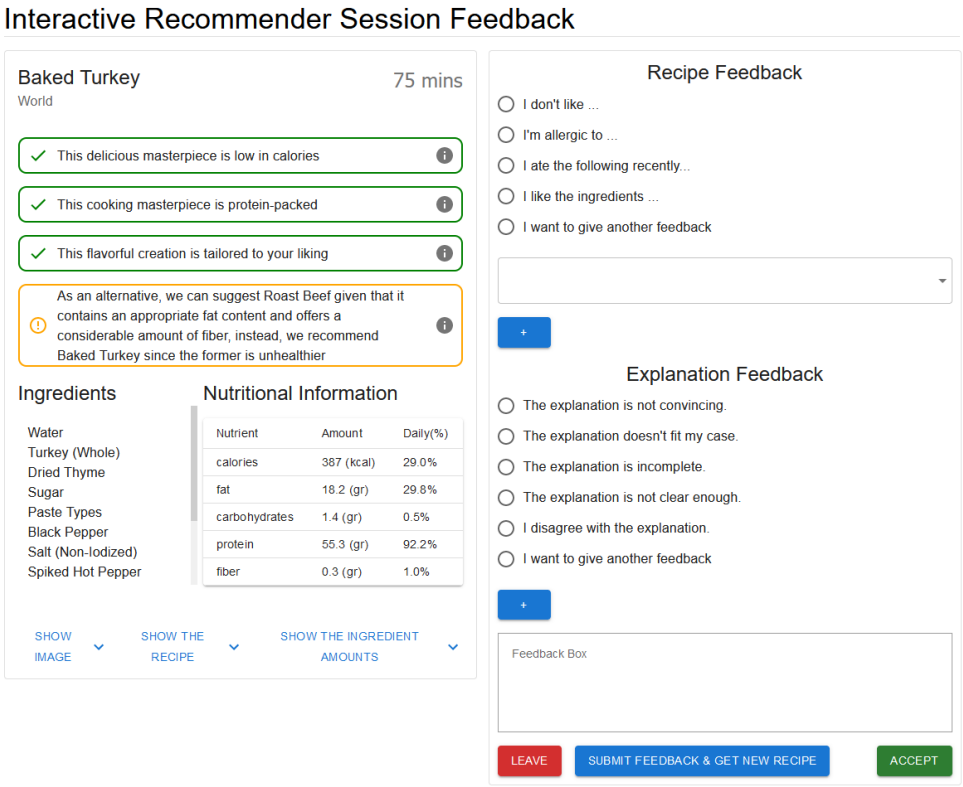

Figure 5.5 – Screenshot of the new interactive system.

The OZU team ran additional studies with the new system we realized with 54 participants which reveal a promising outcome for the explanation generation mechanisms. A thorough analysis of the results and the experiment participants in [wp5-2]

[T5.5 - (Planned to start at M29] Experimentation and validation:
The HES-SO, UNILU, and OZU teams are designing the experimental protocol to evaluate and validate the proof-of-concept system in multi-cultural, multi-national, and inter-continental populations to cover the broadest possible range of preferences, cultural and environmental factors.

References WP5

[wp5-1] D. Calvaresi, R. Carli, J.-G. Piguet, V. H. Contreras, G. Luzzani, A. Najjar, J.-P. Calbimonte, and M. Schumacher. Ethical and legal considerations for nutrition virtual coaches. AI and Ethics, 1–28, 2022.
[wp5-2] Buzcu, B., Varadhajaran, V., Tchappi, I., Najjar, A., Calvaresi, D., Aydoğan, R. (2023). Explanation-Based Negotiation Protocol for Nutrition Virtual Coaching. PRIMA 2022. Lecture Notes in Computer Science, vol 13753. Springer, Cham. https://doi.org/10.1007/978-3-031-21203-1_2

chist-era

1.2. Transnational collaboration

Describe the added value and synergies in the collaboration, any obstacles to the transnational collaboration, and the proposed solution (if necessary).

Concerning WP1:
The HES-SO, OZU, UNIBO, LIST, and UNILU teams attended the monthly steering meetings led by the project coordinator. In particular, all the partners have presented their periodical advancements, which have been punctually discussed with the partners, fostering know-how sharing, obtaining feedback, and implementing new unplanned collaborations to guarantee the task's success and actualize emerging opportunities.

Concerning WP2:
The collaboration among the UNIBO and HES-SO team has mostly regarded task T2.3. Here, interactions concerned eliciting mutual requirements w.r.t. semantic knowledge extraction and manipulation, which should, in turn, support a smooth integration of the software tools developed as part of task T2.4 with the any other software tool to be developed as part of WP3–5.

OZU team worked in collaboration with UNILU to publicize a paper on trust metrics of explainable recommender systems (Task 2.5). To this end, online meetings based on the need were organized usually each Friday from 2PM to 3PM.

Two full working days were organized and co-located with the PRIMA conference in Valencia, Spain. All the partners presented their advancements and used the available time to finalize the online collaborations. Moreover, all the partners have elaborated on possible follow-ups of the developed technologies.

Concerning WP3:
OZU team has worked with UNIBO, LIST, and HES-SO to develop explainability for explainable recommender systems using ML-based methods via symbolic knowledge extraction. Additionally, the OZU team has continued their collaboration with UNILU to improve the explanation generation by the system and established new meetings with UNIBO for the furtherment of the designed human-agent recommendation protocol.

Concerning WP4:
The HES-SO and OZU teams had weekly meetings on WP4 about user profiling aspects, which were beneficial for aligning the tasks T4.1 and T4.3.
The OZU, LIST, and the UNILU teams, as well as HES-SO and UNILU, have worked closely with weekly or bi-weekly meetings to discuss and generally develop the negotiation protocols as well as the initial strategies for the human agent negotiation.
The OZU, HES-SO, and UNIBO teams met (based on the need) to plan/work on the argumentation-based negotiation survey.

Concerning WP5:
The collaboration among the UNIBO and HES-SO team has mostly regarded task T5.1. Here, the interactions concerned eliciting mutual requirements w.r.t. WP5's prototype as well as its conceptual and software architecture. It is worth highlighting that first year's tasks were mostly conceived as self-contained tasks, where the degree of required interaction was low, to let the teams work in parallel as much as possible. Such effort has to be merged and integrated in the second and third year, hence we expect interactions to progressively increase.

The HES-SO and OZU teams had meeting (based on the need) regarding WP2/5 - design of a collection of food ontologies that would feed the overall system.

The HES-SO, LIST, and UNILU teams had (based on the need) meetings involving law & AI Ph.D. students visiting UNILU. Such a collaboration generated a paper (to appear in EXTRAAMAS 2022) and the delineation of legal boundaries for NVC.
The HES-SO and UNILU teams met in the occasion of BNAIC 2021 (conference) to present their joint papers and work InSite in the context of WP5.
OZU team has started working with UNIBO team to bring the multi agent system to life using Python, SPADE library, with weekly meetings.

UNILU (Igor Tchappi) and LIST (Amro Najjar) visited the OZU team in September 2022 (one week) in Istanbul. The goal was to work jointly on the robot implementation and discuss the interface and the user experiments. UNILU (Igor Tchappi) and LIST (Amro Najjar) visited June 2022 (4 days) the HES-SO team. The robot implementation was discussed as well as the metrics of explanation and the questionnaire paper.

Difficulties:
Due to the COVID-19 pandemic, all interactions among teams have occurred online (videoconference). Although this situation might have affected the collaboration negatively, we profited even more from such form of collaboration. Indeed, it allowed the teams to perform more fine-grained and fast-paced interactions. The remarkable increase of online meeting has smoothly blended within the partners schedules. Nevertheless, we restarted in-person events starting from June 2022.

1.3. Significant events and results

(Indicative length: 2-4 pages)

Describe the main achievements of the project. For example:

- *New ideas, new knowledge, new interpretative models of complex phenomena;*
- *Realization of new scientific instrumentation and/or advanced devices;*
- *Implementation of new advanced scientific methodologies;*
- *Realization of prototypes;*
- *Proposal of new technologies;*
- *Contribution to innovation in the production of goods and services;*
- *Development of innovative software;*
- *Economic impact and results exploitation.*

For each achievement, provide a description with factual and, if relevant, quantitative information.

For significant results you would like to publicize using the communication channels of CHIST-ERA, please feel free to forward the information to CHIST-ERA Joint Secretariat using the Toolbox dedicated to the funded projects: https://www.chistera.eu/toolbox.

EREBOTS 2.0
EREBOTS 2.0 is an agent-based GDPR-compliant chatbot platform to deploy 1-to-1 (user-agent) assistive chatbots. Such chatbots (namely agents) can interact based on structured interactions and actualized (zero-code) behavioral change strategies suggested by professional nutritionists. We are currently extending the agents' capabilities, getting the platform ready to allocate the other partners' libraries, and enhancing the interaction methods (i.e., moving towards NLP for selected interactions).

An active Learning approach with the orientation of explainability is deployed in the interactive learning framework for user preference elicitation.

This design has a novel information acquisition strategy for Active Learning through explanations that users can understand and respond to. Besides, the OZU team has currently worked on developing novel uncertainty quantification and information incorporation into ML techniques for Active Learning.

A novel human-agent protocol tailored to food recommendation applications allows an agent to generate both offers and explain why those offers are made and enable the users to give feedback on both given offers and explanations.
The collaboration between UNILU and OZU produced the prototype of a Web application incorporating the designed interaction protocol where the user can specify their constraints and feedback in a structured way without requiring expression in natural language. In particular, the web app allows an agent to (i) generate both offers and explain why those offers are made and (ii) allow the users to give feedback on the given offers and criticize or comment on the explanations provided.

The idea of SKE and SKI as general mechanisms for XAI – coming with several possible methods handling particular situations (e.g., neural networks/support vector machines, classifiers/regressors, discrete/continuous data, etc.) – is indeed novel. Currently, SKE/SKI-related works from the literature focus on proposing particular methods for SKE/SKI or improving existing methods. Poor care is devoted to making these contributions interchangeable, comparable, or interoperable. The conceptual and technological frameworks developed in WP2 serve precisely these purposes. With respect to XAI-related research, the software tools developed in T2.4 (namely, PSyKE and PSyKI from Section 1.1) can be simultaneously described as (i) new scientific instrumentation, (ii) prototypes of general-purpose technologies, and (iii) innovative software toolkits. In fact, to the best of our knowledge, analogous or similar software tools supporting either SKE or SKI were missing. This hindered the potential of XAI-related research by hardening experiments in this field and represented a challenge for the EXPECTATION project. For all these reasons, we argue the proposed software tools have the potential to outlive the project, other than being fundamental to the development of WP3-5.

The HES-SO and UNILU effort in the field of nutrition virtual coaches' (NVC) ethics and legal boundaries expanded and upgraded the existing literature on the ethics of recommender systems [1] and the ethics of food recommender systems [2]. In particular, a tech report (evolving into a journal paper) defines the main ethical and legal challenges and implications for NVC.
[1] Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. Ai & Society, 35, 957-967.
[2] Karpati, D., Najjar, A., & Ambrossio, D. A. (2020, February). Ethics of food recommender applications. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 313-319).

1.4. Technology readiness level (TRL)

Describe the global positioning of the project (from 'idea to application' or from 'lab to market'). Refer to Technology Readiness Levels (see definition here) at the beginning and the end of the project.

- As of now, EREBOTS 2.0 is advancing from TRL3 (current level) to TRL4.
- The implementation of active learning for preference elicitation realized by OZU is between TRL2 and TRL3. It is not expected to evolve further since its mechanisms will be integrated into the EREBOTS 2.0 platform.
- The interaction protocol is TRL2.
- As of now, SKE-SKI have TRL3, and we expect them to reach TRL4 by the end of the project. The tools are being developed following sound software engineering practices, including, but not limited to, distributed version control, unit tests, and continuous integration—guaranteeing they can work on most common operative systems (Windows, MacOS, and Linux).

1.5. Consortium meetings

Provide the cumulative list of consortium meetings from project start.

Meetings				
N°	**Date**	**Location**	**Attending partners**	**Purpose**
1.1 – 1.23	Monthly SC meeting (each first Friday of the month)	Virtual	All partners	Steering Committee meeting to monitor and update on technical advance
2.1 – 2.n	Weekly [M5– M12]	Virtual	UNILU – OZU - LIST	WP4
3.1 –3.n	(bi)weekly [M5 - M12]	Virtual	HES-SO - OZU	WP4 & WP5 & WP2
4.1 –4.n	(bi)weekly [M1 – M12]	Virtual	HES-SO - UNIBO	WP2 & WP5
5	10-14/11/2021	Luxembourg	HES-SO – UNILU - LIST	Attending BNAIC21 and working on T4.2 and WP5 (T5.2)
6.1 –6.n	(bi)weekly [M2 – M10]	Virtual	HES-SO – UNILU - LIST	WP5 (T5.2)
7	17/11/2021	Virtual	ALL	General Assembly
8.1 –8.n	[M1 – M5]	Virtual	UNIBO – OZU -HES-SO	WP5 (T5.1)
1	8-10/06/2022	Bologna	ALL	General Assembly
1	13 – 19/11/2022	Valencia, Spain	All partners	General Assembly
1	2-6/05/2023	Sierre, Switzerland	All partners	General Assembly
1	9 – 14/10/2023	Istanbul, Turkey	All partners	General Assembly

1.6. Deliverables

Provide the cumulative list of deliverables from project start.

Deliverables					
N°	**Title**	**Nature**	**Delivery date (month)**		**Partner in charge**
			Contractual	**Actual**	
Deliverables link (for the CHIST-ERA Commission): https://expectation.ehealth.hevs.ch/posts/deliverables/					
D1.1	Yearly Report	Report	M12	M12	HES-SO/ALL
D1.4	Data Management Plan	Document	M3	M3	HES-SO
D2.1	Tech report on symbolic knowledge extraction/injection	Tech Report	M8	M12*	UNIBO
D2.2	Scientific paper on symbolic knowledge extraction/injection	Paper	M12	M13* (under review)	UNIBO
D2.3	Reusable Software library for intra-agent explainability	Software	M14	M14	UNIBO
D3.1	Technical report detailing the developed models and data integration	Tech report	M25	M27	OZU
D3.2	Scientific paper focusing on heterogeneous data integration and conflict resolution.	Paper	M27	n/a	OZU
D3.2a	Scientific paper focusing on heterogeneous data integration	Paper	M27	n/a	OZU
D3.2b	Scientific paper focusing on conflict resolution	Paper	M27	n/a	OZU

D4.1	Technical report detailing the developed user models and agent-based profiling	Tech report	M20	M25	HES-SO
D5.1	Technical report detailing architectural, ethical and societical reqirements	Tech report	M16	M20	HES-SO

* See explanations of delays in section 1.1 (project objectives and activities implemented) and 3.1 (comments on expenses)

1.7. Free comments

Compliance with project objectives, interaction between the partners, issues, questions to CHIST-ERA...

To request a project modification, please use the dedicated form on the Toolbox: https://www.chistera.eu/toolbox

Dr. Amro Najjar moved from the University of Luxembourg (UNILU) to the Luxembourg Institute of Science and Technology (LIST) where he is now working on a permanent position and a full-time researcher. Since Dr. Najjar was the main contributor on the UNILU side, we requested that he continues to work on the project with his new affiliation at LIST. In addition, to complete the person-month freed by Dr. Najjar, Igor Tchappi has been hired as a UNILU postdoc. He started to work on the project from the 15th of February 2022. This amendment would not imply an increase of the budget, the university would transfer to LIST the personnel costs of Dr. Najjar's contribution. We discussed this with the FNR (the national funding agency), and with all of the project partners --- who unanimously agreed. The coordinator (HES-SO) is taking the contractual steps (update of DoA, CA and change request).

chist-era

2. Dissemination of results, exploitation, impact

2.1. Scientific publications (conferences/workshops, book chapters, etc.)

Indicate the publications resulting from the project. Mention only those that result directly from the project (after it started, and which mention the support of CHIST-ERA and the project reference). Indicate whether they correspond to single or multi-partner communications (multi-partner means involving several project partners). Indicate if they are available in Open Access and linked to the respective underlying data. Provide the corresponding Digital Object Identifiers (DOI).

Distinguish the different categories of publications (journals/conference proceedings, technical reports, etc.). Use the usual citation standards for the field reference. If the publication is accessible on line, indicate the URL.

Please harmonise the bibliography and use only one font.

	Scientific publications					
Reference (list of authors, journal/conference proceedings/other, pages, year of publication, ...)	**Multi-project partners of same country (Yes/No)**	**Multi-project partners of different countries (Yes/No)**	**Open Access (Yes/No)**	**DOI**	**URL**	**DOI(s) of underlying data**
Federico Sabbatini, Giovanni Ciatto, Roberta Calegari, Andrea Omicini. "On the Design of PSyKE: A Platform for Symbolic Knowledge Extraction". WOA 2021 – 22nd Workshop "From Objects to Agents". CEUR Workshop Proceedings 2963, October 2021	**No**	**No**	**Yes**	**none**	http://ceur-ws.org/Vol-2963/paper14.pdf	
Giuseppe Pisano, Roberta Calegari, Andrea Omicini. "Towards cooperative argumentation for MAS: An actor-based approach". WOA 2021 – 22nd Workshop "From Objects to Agents". CEUR Workshop Proceedings 2963, October 2021	no	no	yes	none	TBD	
Andrea Agiollo, Giovanni Ciatto, Andrea Omicini. "Shallow2Deep: Restraining Neural Networks Opacity through Neural Architecture Search". Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021	no	no	no	10.1007/978-3-030-82017-6_5	http://ceur-ws.org/Vol-2963/paper17.pdf	
Federico Sabbatini, Giovanni Ciatto, Andrea Omicini. "GridEx: An Algorithm for Knowledge Extraction from Black-Box Regressors". Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021	no	no	no	10.1007/978-3-030-82017-6_2	https://doi.org/10.1007/978-3-030-82017-6_5	
Giovanni Ciatto, Amro Najjar, Jean-Paul Calbimonte, Davide Calvaresi. "Towards Explainable Visionary Agents: License to Dare and Imagine". Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021	no	yes	no	10.1007/978-3-030-82017-6_9	http://dx.doi.org/10.1007/978-3-030-82017-6_2	

Davide Calvaresi, Giovanni Ciatto, Amro Najjar, Reyhan Aydoğan, Leon Van der Torre, Andrea Omicini, Michael I. Schumacher. "Expectation: Personalized Explainable Artificial Intelligence for Decentralized Agents with Heterogeneous Knowledge". Explainable and Transparent AI and Multi-Agent Systems. Third International Workshop, EXTRAAMAS 2021	no	yes	no	10.1007/978-3-030-82017-6_20	http://dx.doi.org/10.1007/978-3-030-82017-6_9	
Andrea Agiollo, Giovanni Ciatto, Andrea Omicini. "Graph Neural Networks as the Copula Mundi between Logic and Machine Learning: A Roadmap". WOA 2021 – 22nd Workshop "From Objects to Agents". CEUR Workshop Proceedings 2963, October 2021	no	no	yes	no	http://dx.doi.org/10.1007/978-3-030-82017-6_20	
Yazan Mualla, Igor Tchappi, Timotheus Kampik, Amro Najjar, Davide Calvaresi, Abdeljalil Abbas-Turki, Stéphane Galland, Christophe Nicolle: The quest of parsimonious XAI: A human-agent architecture for explanation formulation. Artif. Intell. 302: 103573 (2022)	No	Yes	yes	https://doi.org/10.1016/j.artint.2021.103573	http://ceur-ws.org/Vol-2963/paper18.pdf	
Rachele Carli, Amro Najjar: Rethinking Trust in Social Robotics. CoRR abs/2109.06800 (2021)	No	Yes	Yes	https://doi.org/10.48550/arXiv.2109.06800	(link)	
Contreras, V., Aydogan, R., Najjar, A., & Calvaresi, D. On Explainable Negotiations via Argumentation	No	Yes	No	No	https://luis.leiva.name/tmp/bnaic2021_preproceedings.pdf	
Giovanni Ciatto, Roberta Calegari, Andrea Omicini. "2P-Kt: A Logic-Based Ecosystem for Symbolic AI". SoftwareX 16, December 2021	Yes	No	Yes	https://doi.org/10.1016/j.softx.2021.100817	https://www.sciencedirect.com/science/article/pii/S2352711021001126	
Buzcu, B., Varadhajaran, V., Tchappi, I., Najjar, A., Calvaresi, D., Aydoğan, R. (2023). Explanation-Based Negotiation Protocol for Nutrition Virtual Coaching. PRIMA 2022. Lecture Notes in Computer Science(), vol 13753. Springer, Cham.	Yes	Yes	No	https://doi.org/10.1007/978-3-031-21203-1_2	https://link.springer.com/chapter/10.1007/978-3-031-21203-1_2	
Furkan Cantürk, Reyhan Aydoğan. Explainable Active Learning for Preference Elicitation, 29 August 2023,	No	No	Yes	https://doi.org/10.212	https://www.researchsquare.com/article/	

				03/rs.3.rs-3295326/v1	rs-3295326/v1	
Ciatto, G., Magnini, M., Buzcu, B., Aydoğan, R., Omicini, A. (2023). A General-Purpose Protocol for Multi-agent Based Explanations. Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2023. Lecture Notes in Computer Science vol 14127. Springer, Cham.	Yes	Yes	No	http://dx.doi.org/10.1007/978-3-031-40878-6_3	https://link.springer.com/chapter/10.1007/978-3-031-40878-6_3	
Hulstijn, J., Tchappi, I., Najjar, A., Aydoğan, R. (2023). Metrics for Evaluating Explainable Recommender Systems. Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2023. Lecture Notes in Computer Science(), vol 14127. Springer, Cham.	Yes	Yes	No	https://doi.org/10.1007/978-3-031-40878-6_12	https://link.springer.com/chapter/10.1007/978-3-031-40878-6_12	
Berk Buzcu, Melissa Tessa, Igor Tchappi et al. Towards Interactive Explanation-based Nutrition Virtual Coaching Systems, 29 June 2023, PREPRINT (Version 1) available at Research Square	Yes	Yes	Yes	https://doi.org/10.21203/rs.3.rs-3110083/v1	https://www.researchsquare.com/article/rs-3110083/v1	
Sabbatini, Federico et al. 'Symbolic Knowledge Extraction from Opaque ML Predictors in PSyKE: Platform Design & Experiments'. 1 Jan. 2022 : 27 – 48.	No	No	No	10.3233/IA-210120	https://content.iospress.com/articles/intelligenza-artificiale/ia210120	
Magnini, M., Ciatto, G., Omicini, A. (2022). On the Design of PSyKI: A Platform for Symbolic Knowledge Injection into Sub-symbolic Predictors. In: Calvaresi, D., Najjar, A., Winikoff, M., Främling, K. (eds) Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2022. Lecture Notes in Computer Science(), vol 13283. Springer, Cham	No	No	No	10.1007/978-3-031-15565-9_6	https://link.springer.com/chapter/10.1007/978	
Federico Sabbatini, Giovanni Ciatto, Roberta Calegari, Andrea Omicini. Hypercube-Based Methods for Symbolic Knowledge Extraction: Towards a Unified Model. In: Proceesings of the 23rd Workshop "From Objects to Agents" (WOA 2022). Ceur Workshop Proceedings	No	No	Yes		http://ceur-ws.org/Vol-3261/paper4.pdf	

chist-era

Sabbatini, F., Ciatto, G., Omicini, A. (2022). Semantic Web-Based Interoperability for Intelligent Agents with PSyKE. In: Calvaresi, D., Najjar, A., Winikoff, M., Främling, K. (eds) Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2022. Lecture Notes in Computer Science(), vol 13283. Springer, Cham	No	No	Yes	10.1007/978-3-031-15565-9_8	https://link.springer.com/chapter/10.1007/978-3-031-15565-9_8	
Matteo Magnini, Giovanni Ciatto, Andrea Omicini. A view to a KILL: Knowledge Injection via Lambda Layer. In: Proceesings of the 23rd Workshop "From Objects to Agents" (WOA 2022). Ceur Workshop Proceedings	No	No	Yes		http://ceur-ws.org/Vol-3261/paper5.pdf	
Matteo Magnini, Giovanni Ciatto, Andrea Omicini. KINS: Knowledge Injection via Network Structuring. In: Proceesings of the 37th Italian Conference on Computational Logic (CILC 2022). Ceur Workshop Proceedings	No	No	Yes	http://ceur-ws.org/Vol-3204/paper_25.pdf		
Ciatto, G., Magnini, M., Buzcu, B., Aydoğan, R., Omicini, A. (2023). A General-Purpose Protocol for Multi-agent Based Explanations. In: Calvaresi, D., *et al.* Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2023. Lecture Notes in Computer Science(), vol 14127. Springer, Cham.	No	No	Yes	https://doi.org/10.1007/978-3-031-40878-6	https://link.springer.com/book/10.1007/978-3-031-40878-6	
Andrea Rafanelli, Stefania Costantini, Andrea Omicini. Position Paper: On the Role of Abductive Reasoning in Semantic Image Segmentation. In: Proceedings of the 21st International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022). Ceur Workshop Proceedings	No	No	Yes		https://ceur-ws.org/Vol-3419/paper9.pdf	
Agiollo, A., Cavalcante Siebert, L., Murukannaiah, P.K., Omicini, A. (2023). The Quarrel of Local Post-hoc Explainers for Moral Values Classification in Natural Language Processing. In: Calvaresi, D., *et al.* Explainable and Transparent AI and Multi-Agent Systems. EXTRAAMAS 2023. Lecture Notes in Computer Science(), vol 14127. Springer, Cham.	Yes	No	No	10.1007/978-3-031-40878-6_6	https://link.springer.com/chapter/10.1007/978-3-031-40878-6_6	
Sabbatini, Federico et al. 'Towards a Unified Model for Symbolic Knowledge Extraction with Hypercube-based Methods'. 1 Jan. 2023 : 63 – 75.	No	No	No	10.3233/IA-230001	https://content.iospress.com/articles/intelligenza-	

chist-era

					artificiale/ia230001	
Matteo Magnini, Giovanni Ciatto, Furkan Cantürk, Reyhan Aydoğan, Andrea Omicini. Symbolic knowledge extraction for explainable nutritional recommenders. In: Computer Methods and Programs in Biomedicine, Volume 235, 2023, 107536, ISSN 0169-2607	No	No	No	10.1016/j.cmpb.2023.107536	https://pubmed.ncbi.nlm.nih.gov/37060685/	
Alcaraz, B., Hosseini-Kivanani, N., Najjar, A., & Bongard-Blanchy, K. (2023, March). User Requirement Analysis for a Real-Time NLP-Based Open Information Retrieval Meeting Assistant. In *European Conference on Information Retrieval* (pp. 18-32). Cham: Springer Nature Switzerland.	Yes	No	No		https://link.springer.com/chapter/10.1007/978-3-031-28244-7_2	
Alcaraz, B., Hosseini-Kivanani, N., & Najjar, A. (2022, July). IRRMA: An Image Recommender Robot Meeting Assistant. In International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 449-453). Cham: Springer International Publishing	Yes	No	No	10.1007/978-3-031-18192-4_36	https://link.springer.com/chapter/10.1007/978-3-031-18192-4_36	
Carli, R., Najjar, A., & Calvaresi, D. (2022, December). Human-Social Robots Interaction: The blurred line between necessary anthropomorphization and manipulation. In Proceedings of the 10th International Conference on Human-Agent Interaction (pp. 321-323).	No	Yes	No	10.1145/3527188.3563941	https://dl.acm.org/doi/10.1145/3527188.3563941	
Carli, R., Najjar, A., & Calvaresi, D. (2022, May). Risk and Exposure of XAI in Persuasion and Argumentation: The case of Manipulation. In International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems (pp. 204-220). Cham: Springer International Publishing.	No	Yes	No	10.1007/978-3-031-15565-9_13	https://dl.acm.org/doi/10.1007/978-3-031-15565-9_13	
Davide Calvaresi, Rachele Carli, Jean-Gabriel Piguet, Victor H. Contreras, Gloria Luzzani, Amro Najjar, Jean-Paul Calbimonte & Michael Schumacher. Ethical and legal considerations for nutrition virtual coaches. AI Ethics, 1–28 (2022).	No	Yes	Yes	10.1007/s43681-022-00237-6	https://link.springer.com/article/10.1007/s43681-022-00237-6	
Tchappi, I., Mboula, J. E. N., Najjar, A., Mualla, Y., & Galland, S. (2022). A decentralized multilevel agent based explainable model for fleet management of remote drones. Procedia Computer Science, 203, 181-188.	Yes	No	Yes	10.1016/j.procs.2022.07.025	https://www.sciencedirect.com/science/article/pii/S18	

					77050922006305	
Tchappi, I., Mualla, Y., Galland, S., Bottaro, A., Kamla, V. C., & Kamgang, J. C. (2022). Multilevel and holonic model for dynamic holarchy management: Application to large-scale road traffic. Engineering Applications of Artificial Intelligence, 109, 104622.	No	No	No	10.1016/j.engappai.2021.104622	https://www.sciencedirect.com/science/article/abs/pii/S0952197621004358	
Calvaresi, D., Najjar, A., Winikoff, M., & Främling, K. (Eds.). (2022). Explainable and Transparent AI and Multi-Agent Systems: 4th International Workshop, EXTRAAMAS 2022, Virtual Event, May 9–10, 2022, Revised Selected Papers (Vol. 13283). Springer Nature.	Yes	Yes	No	10.1007/978-3-031-15565-9	https://link.springer.com/book/10.1007/978-3-031-15565-9	
Contreras, V., Schumacher, M., & Calvaresi, D. (2022, May). Integration of local and global features explanation with global rules extraction and generation tools. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems* (pp. 19-37). Cham: Springer International Publishing.	No	No	Yes		https://arodes.hes-so.ch/record/11423/files/Contreras_2022_integration_local_global.pdf	
Contreras, V., Marini, N., Fanda, L., Manzo, G., Mualla, Y., Calbimonte, J. P., ... & Calvaresi, D. (2022). A DEXiRE for extracting propositional rules from neural networks via binarization. *Electronics*, *11*(24), 4171.	Yes	No	Yes		https://www.mdpi.com/2079-9292/11/24/4171	
Contreras, V., Bagante, A., Marini, N., Schumacher, M., Andrearczyk, V., & Calvaresi, D. (2023, May). Explanation Generation via Decompositional Rules Extraction for Head and Neck Cancer Classification. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems* (pp. 187-211). Cham: Springer Nature Switzerland.	Yes	No	Yes		https://link.springer.com/chapter/10.1007/978-3-031-40878-6_11	

URL of Data Management Plan (optional): D1.4-[M3]-DMP_195530-2021.pdf

chist-era

2.2 Exploitation plan

Outline an exploitation plan of your most significant exploitable results including:

- *Who will exploit the result output (project participant/if someone else then who and how will they be informed);*
- *Use type (commercial/other use);*
- *Intellectual property rights arrangements if relevant;*
- *Target end user;*
- *Roadmap and goals during and after the project's lifetime (plan of actions to be taken to achieve exploitation);*
- *Timeframe.*

Use type, intellectual properties, and Licensing:

The **software tools produced in WP2** are freely available on the Web under an open-source license (Apache 2p) supporting both research and commercial exploitation. Contributions – in the form of bugfixes, feature requests, of original SKE/SKI algorithms implementations – are not only welcome but also encouraged via the GitHub Social Coding platform. Keeping the codebase open and publicly available is part of a deliberate strategy aimed at attracting researchers interested in experimenting with SKE and SKI.

Ad-hoc lectures concerning SKE and SKI have been scheduled as part of the "Intelligent System Engineering" and "Multi Agent Systems" courses held at University of Bologna as part of the Master's degree programmes in "Computer Science and Engineering" and "Artificial Intelligence", respectively.

EREBOTS 2.0 will be released at the end of the project with an open-source license (under evaluation – possibly BSD 3 clause).
While the basic versions of the software will always be kept free and open source, we expect to evaluate the possibility of producing a "premium" version of the software by the end of the project (in collaboration with possibly interested industrial partners).

Targeted Users:

The recipient(s) of our tools/software are academia, profiting from the algorithmic and implemented innovations, end-users (a selection of the stakeholders identified during this project), and professional nutritionists using the nutrition virtual coach system

RoadMap & Goals (post-project):

- G1) Applying the acquired knowledge to other research challenges.
- G2) Writing other research proposals involving some current consortium partners (i.e., applying the obtained knowledge and actualized prototypes to tackle intra-/inter-agent explainability in other unexplored domains (i.e., neuroscience, sport, and energy domains).
- G3) Connecting with industrial partners (e.g., knowledge transfer via InnoSwiss projects).
- G4) Connecting with medical partners (e.g., conducting further investigations focusing on specific user groups and to fully assess the legal boundaries of intelligent assistive systems in real-world settings).
- G5) Investigating more robot-human interaction (RHI) as future assistive tools for instance for children.

chist-era

2.3 Exploitation overview (software, products, spin-offs, etc.)

Use the table below to outline your current progress in the exploitation plan (see previous section): achievements so far and next steps. Fill in the goals foreseen in your plan for every year of your project and 3 subsequent years after the end of your project (column 1) and actual exploited results up to date (column 2).

Period	Planned goals	Actual exploited results
Year 1	- SKE-SKI libraries	SKE-SKI libraries
Year 2	- First integration external libraries into EREBOTS 2.0 (generic prototype) - Additionally, G3	
Year 3	- Full working scenario-related prototype + open-source code - Additionally, G3, G4	
Project end + 1 year	- Writing of new research proposal to employ the generated new knowledge - Additionally, G1, G2, G4, G5	n/a
Project end + 2 years	- G3	n/a
Project end + 3 years		n/a

Describe project spin-off effects, for example:

- *Software and any other prototype;*
- *Standardization actions;*
- *National and international patents, licences, and other elements of intellectual property;*
- *Launching of product or service, new project, contract, etc.;*
- *Development of a new partnership;*
- *Creation of a platform available to a community;*
- *Company creation, spin-off companies, fund-raising.*

--

2.4 Other dissemination of results

Mention any communication actions, including the project website creation and management and the target audience.

Project website: https://expectation.ehealth.hevs.ch/

Project video for general audience: https://expectation.ehealth.hevs.ch/posts/promo/

Round Table (scientific panel) about the EXPECTATION project at EXTRAAMAS 2022

EXTRAAMAS International workshop (CHIST-ERA special track): https://extraamas.ehealth.hevs.ch/

Deliverables: https://expectation.ehealth.hevs.ch/posts/deliverables/

chist-era

3. Resources and Funding

3.1. Project level (from project start) -- M1-M24

Budget used				
N°	**Partner**	**Person months (PM)**	**Total costs (EUR)**	**Percentage of requested budget**
1	HES-SO	33 PM	246.000,00	76%
2	UNIBO	31.22	123.493,2	97%
3**	OZU	45.4	12.594,74	51 %
4	UNILU	16	140.428,88	54%
5	LIST	4.39	41.440,40	47%

Comments on expenses

UNIBO:

The Italian partner (UNIBO) has, to this day, not yet received funds.
Somehow, the PI of UNIBO (Andrea Omicini) has managed to circumvent such a huge problem making use of some internal funds (anticipating the grant's money). This has led to minimal delays into delivering D2.1 and D2.2.
Nevertheless, the Italian partner has very well contained such a circumstance, managing to contribute as expected and not harming anyhow the project. Indeed, the reporting delays have not affected the production of the actual contents.
However, if the founds will not be released shortly, the Italian partner might not be able to sustain the schedule as planned.

OZU:

Given the highly variable exchange rate, the Turkish partner had enormous difficulties when it came to buying material from the European market. In the coming years, they may struggle to arrange their travel expenses for attending the project meetings to be held by other partners, especially when it is held in an expensive country like Switzerland due to the travel budget limitation. The travel budget should be reconsidered, and the CHIST-ERA organization could actualize possible support.

** Total costs refer to the expenses made under the equipment, personnel and bursary, service, travel and other budget lines and actualised as 12.594,74 Euro for the reporting period. The ECB exchange rate for 01.07.2021-31.03.2022 is used for the calculations. (1 euro=29.7198 TL)