A Design Space for Explainable Ranking and Ranking Models

I. Al Hazwani[1,2], J. Schmid[1], M. Sachdeva[1], and J. Bernard[1,2]

[1] University of Zurich, Zürich, Switzerland
[2] Digital Society Initiative, Zürich, Switzerland

Motivation and problem statement

The use of black-box model for generate ranking results do not always reflect a) the ranking order based on inherent item criteria leading to possible bias and b) the preferences of users which lead to not have human-centered ranking. Providing explanations of third-party item rankings and underlying ranking models would bring multiple benefits for different stakeholders namely end-user, data analysis and model developer.

Background

An explainable pipeline is generaly composed of three main blocks[1]:

Explanandum $\to$ Explanator $\to$ Explainee

- **Explanandum** means the complete oracle to be explained.
- **Explanator** is the explanation system that provides the explanation.
- **Explainee** is the receiver of the explanations.
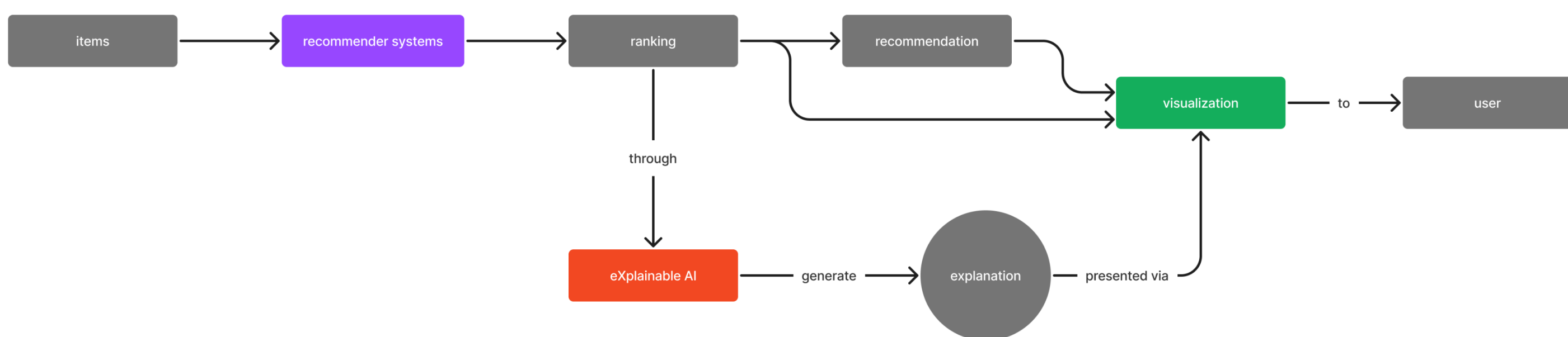
Generating ranking explanation is a cross-domain problem. It requires to use metodologies from:

- **Recommender System** (RS), to find similar items;
- **eXplainable Artificial Intelligence** (XAI), to generate explanation;
- **Visualization** (VIS), to visualized the ranking, reccomendation, and explanation to the user.

User characterization

Based on existing works and the study of real-world ranking applications, we identify three different user groups with considerable differences.
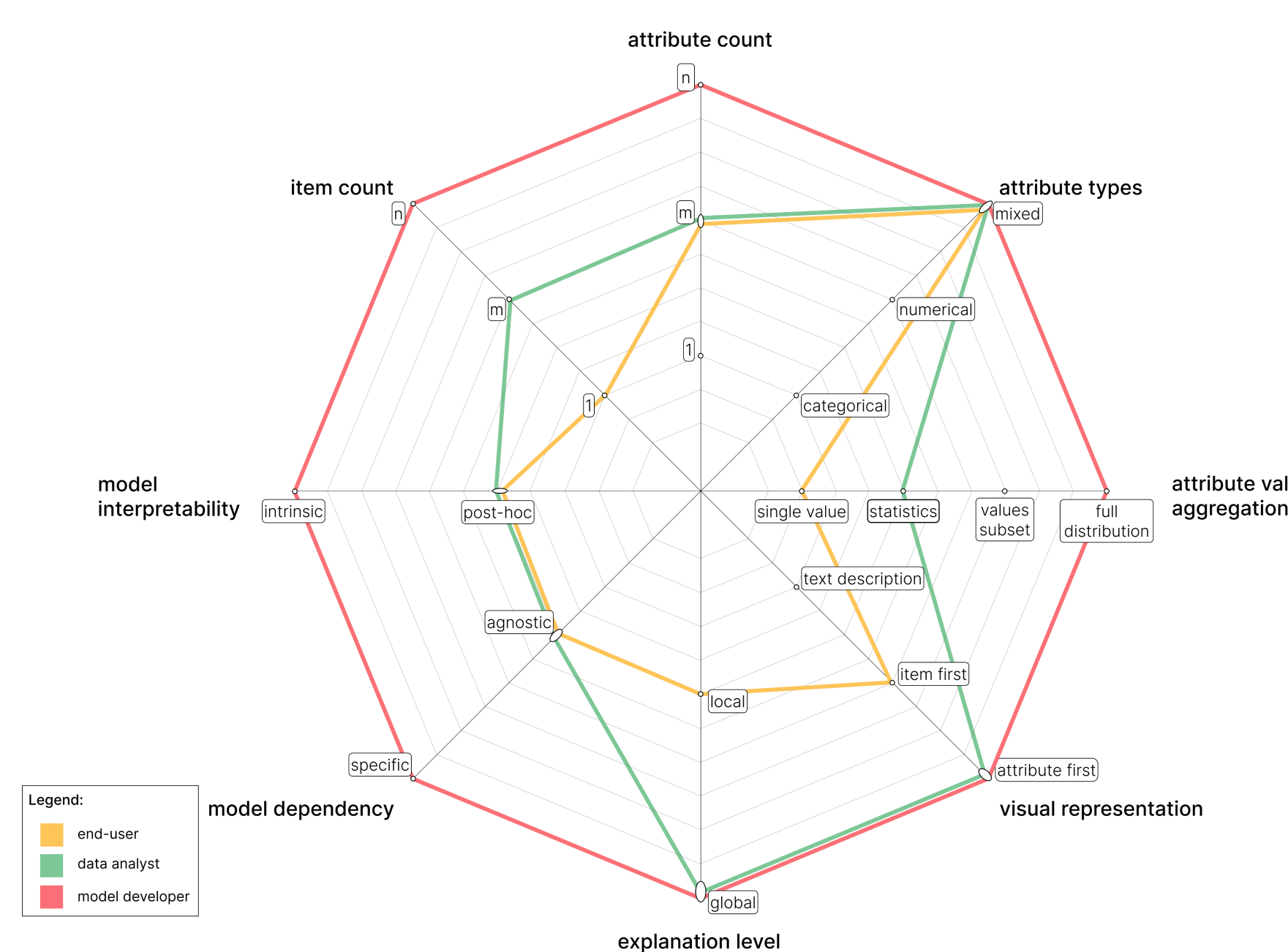
Increasing level of AI knowledge

Model Developers differ from other users by their particular **knowledge on AI** and recommender systems, and by their goal of **model building** as opposed to model-usage.

Data Analysts are interested in the **(visual) analysis** of item rankings. Their main emphasis are on attribute characteristics for large number of items and for **model-based explainability** to draw general conclusion about items, attributes, and model characteristics.

End users want to make decisions based on a given ranking. They aim at **increasing their trust** in the system by **understanding how** the ranking at hand is aligned with their mental model.

Design space

The synthesis of abstractions in XAI, RS, and VIS reveals **eight dominating characteristics** that form the dimensions of the design space. Each dimension contains different levels, always with increasing complexity from the center to the outer, useful to a) describe existing explainers and b) generate uncharted room for yet other types of explainers.
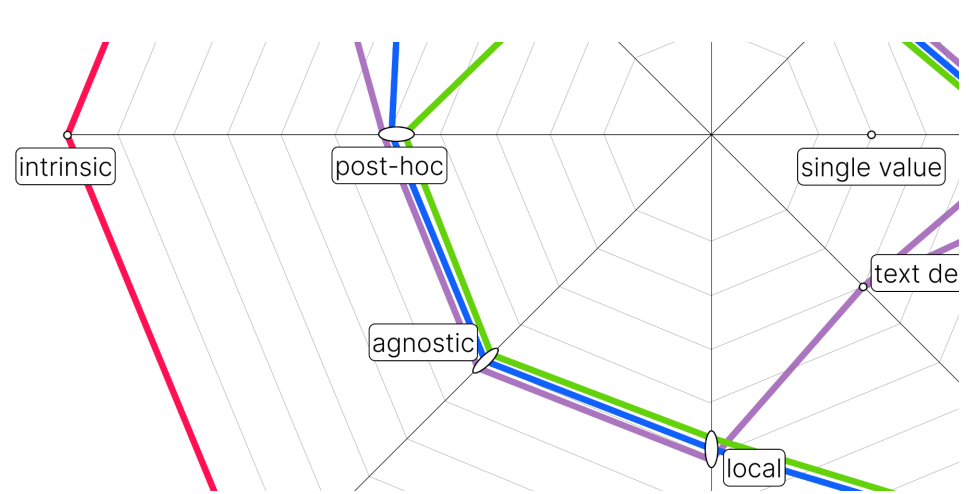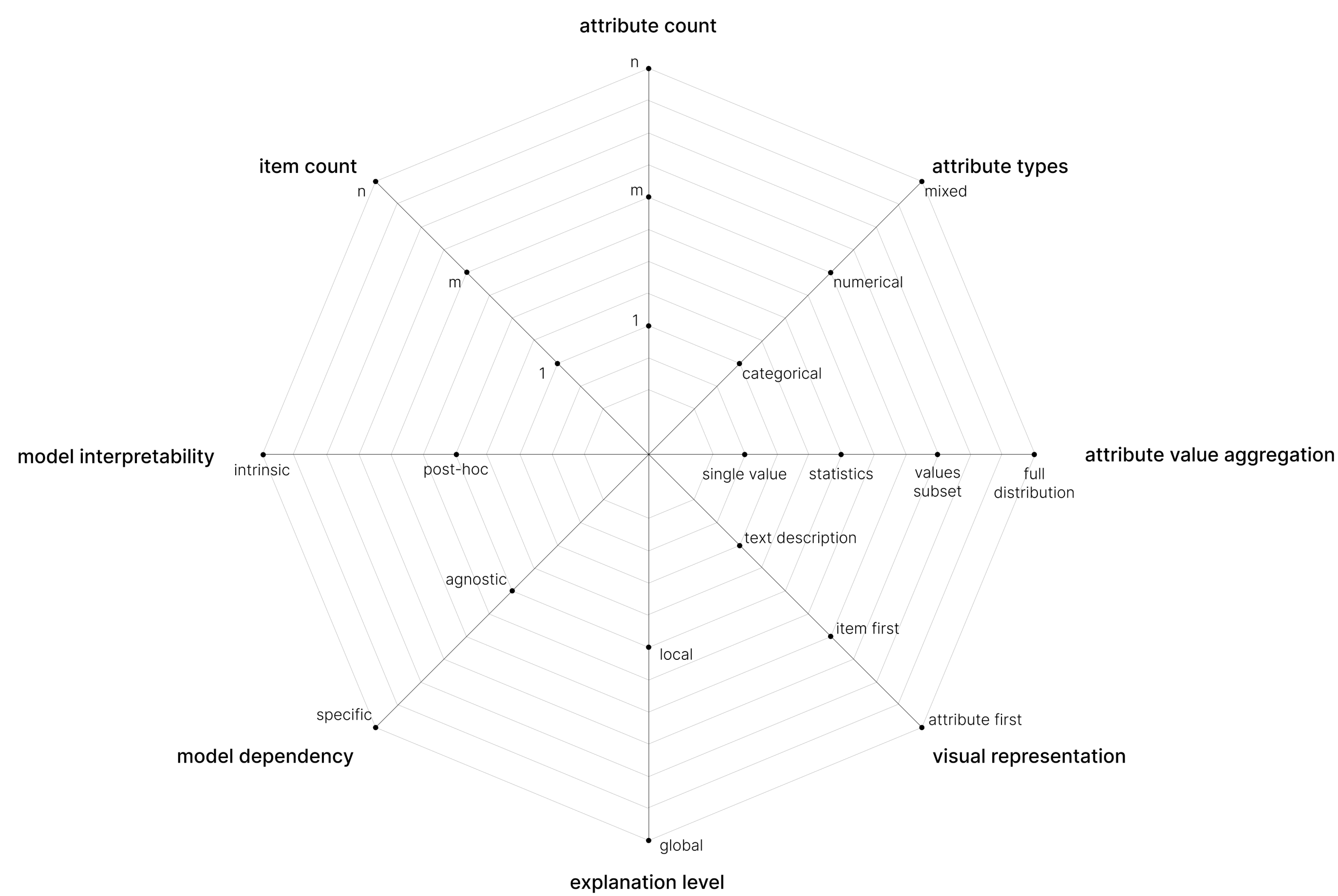
We validate the design space by mapping existing XAI approaches for rankings into the space and user needs, using the notion of radar charts in combination with metro maps[2].

Mapping of user groups

It is essential to take a **human-centered** approach and align the explanations with the user's knowledge and needs. It is possible to leverage the **descriptive power** of the design space to map the three different users' needs to help future designers and developers to create more target-oriented solutions (**generative power** of the design space).

Mapping of existing explainers
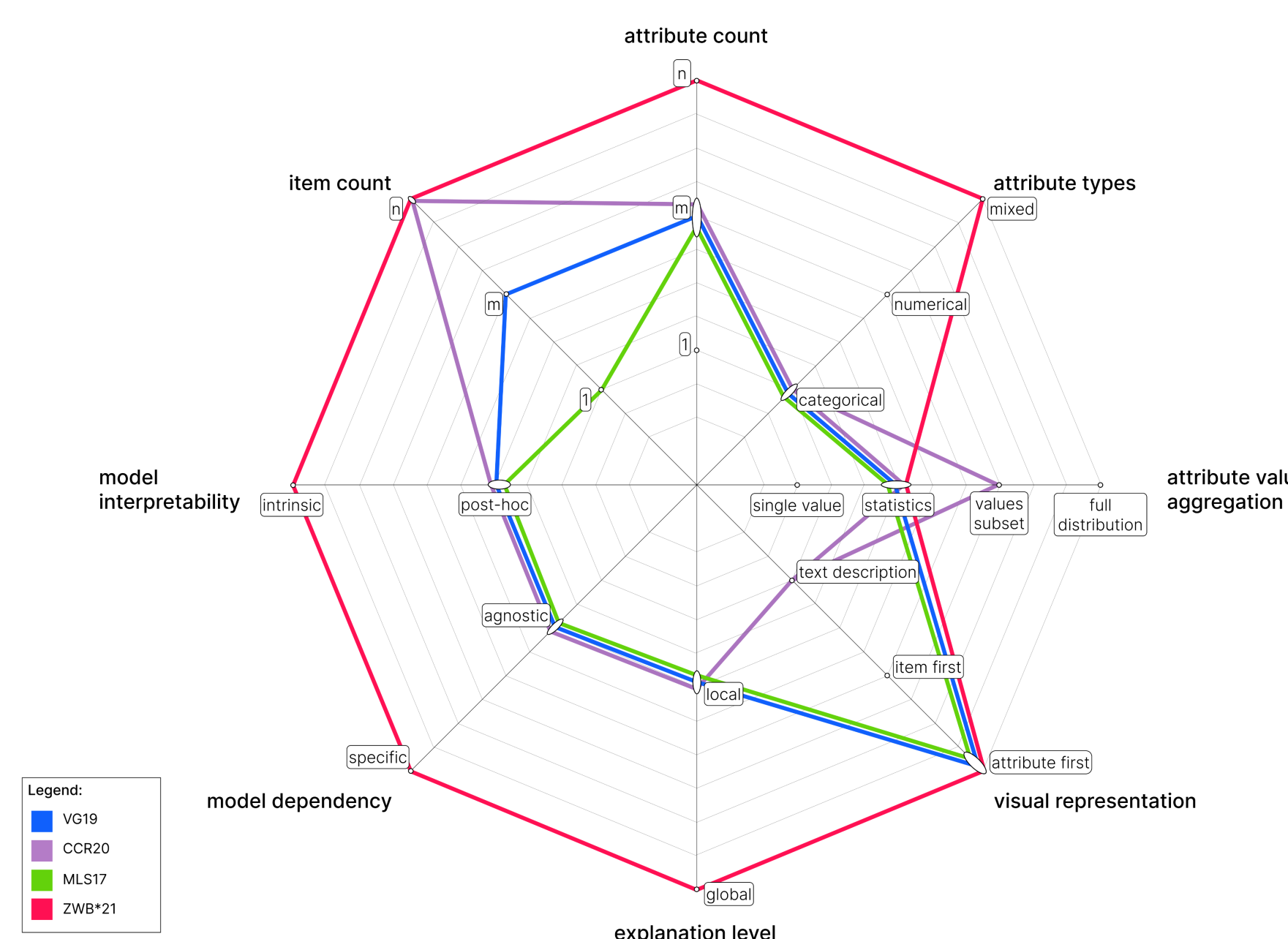
Three of the four XAI approaches[3,4,5] mapped have similar characteristics: local level of explanation, agnostic model dependency, and post-hoc model interpretability. In contrast, one work uses intrinsic model interpretability, a global explanation level, and is by far the most sophisticated XAI solution observed[6].

Future work

Future work on this topic includes: a) expanding the design space to include user interaction, b) leveraging the design space to design a human-centered explainable AI ranking system and c) cross-cutting the design space with industry human-AI guidelines[7].

References

[1] Arrieta, Alejandro Barredo, et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." Information fusion 58: 82-115, 2020.
[2] Neumayer, Robert, et al. "The metro visualisation of component planes for self-organising maps." 2007 International Joint Conference on Neural Networks, 2007.
[3] Verma, Manisha, and Debasis Ganguly. "LIRME: locally interpretable ranking model explanation." Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019.
[4] Choi, Jaekeol, et. al. "Interpreting neural ranking models using grad-cam." arXiv preprint arXiv:2005.05768, 2020.
[5] Muhammad, Khalil, et.al. "On the pros and cons of explanation-based ranking." International Conference on Case-Based Reasoning. Springer, Cham, 2017.
[6] Zhuang, Honglei, et al. "Interpretable Ranking with Generalized Additive Models." Proceedings of the 14th ACM International Conference on Web Search and Data Mining, 2021.
[7] Wright, Austin P., et al. "A comparative analysis of industry human-AI interaction guidelines." arXiv preprint arXiv:2010.11761, 2020.

University of Zurich UZH   IVDA INTERACTIVE VISUAL DATA ANALYSIS   UZH DSI

Contact person: alhazwani@ifi.uzh.ch