"You've Got a Friend in Me": A Formal Understanding of the Critical Friend Agent

Joel Wester
joelw@cs.aau.dk
Aalborg University
Aalborg, Denmark

Andreas Brännström
andreas.brannstrom@umu.se
Umeå University
Umeå, Sweden

Juan Carlos Nieves
juan.carlos.nieves@umu.se
Umeå University
Umeå, Sweden

Niels van Berkel
nielsvanberkel@cs.aau.dk
Aalborg University
Aalborg, Denmark

ABSTRACT

State-of-the-art intelligent and interactive agents, such as Alexa or Siri, often present overly conforming behaviour during interactions with humans. This can result in a misalignment between end-user expectations and agent behaviour. To overcome this barrier in human-AI interactions, we introduce the Critical Friend (CF), a conceptual idea that guides critical behaviour in human-human interactions. We present our results as a formal understanding that can be described through description logic and utilised for reasoning capabilities, enabling implementations of the CF as an intelligent interactive agent.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **HCI theory, concepts and models**; **characterised evaluations**.

KEYWORDS

Critical Friend, agent, formal understanding, characterisation

ACM Reference Format:
Joel Wester, Andreas Brännström, Juan Carlos Nieves, and Niels van Berkel. 2023. "You've Got a Friend in Me": A Formal Understanding of the Critical Friend Agent. In *International Conference on Human-Agent Interaction (HAI '23), December 04–07, 2023, Gothenburg, Sweden.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3623809.3623957

1 INTRODUCTION

People generally have an understanding of what it means to be a friend. Being a good friend in human-to-human interactions requires a careful balance between providing your friend with what they *want*, and what they *need*. Depending on the specifics of the relationship, this balancing act oftentimes flows naturally. In educational research, human-to-human interactions have been conceptualised to guide senior-junior interactions. In such settings, different Critical Friend (CF) frameworks are used to inform how to act in situations which require varying levels of 'critical' and 'friend' [1, 5, 14]. To illustrate, in a situation where a student misbehaves, an increase in 'critical' and a decrease in 'friend'-like behaviour might be necessary to provide the necessary support. Although no rigid understanding of the CF is provided in the literature, Costa et al. conceptually define it to be "*a trusted person who asks provocative questions* [...]" [5, p.50]. Similarly, Storey and Wang describe it as "*a challenging critic and a trusting friend*" [20, p.1]. Therefore, as a response to conforming agents, we seek to better understand how to formalise the Critical Friend (CF) behaviour for a Critical Friend Agent (CFA) and contrast it against different characterisations. We investigate this by exploring how the CF can be formally understood to enable utilisation in human-AI interactions. In this paper, we, therefore, pose the following research question: **How can the Critical Friend Agent be formally understood?**

2 KNOWLEDGE MODELING

The initial step of our approach was to collect available articles online that included a definition of the CF, generate data from these articles, and analyse the data using a Grounded Theory approach [23]. Our formal understanding is therefore based on and informed by prior research from the educational research domain.

We collected online articles that included a definition of the CF. We took inspiration from PRISMA guidelines by structuring our data search and data generation [16]. We limited our search to Google Scholar, including articles that met our four inclusion criteria: time frame (published between 1950-2021), language (English), peer-reviewed in a scientific venue, and 'Critical Friend" in the title. This resulted in 299 articles. After reviewing titles and sorting out irrelevant articles against our inclusion criteria, we included 52 articles. We reviewed the abstracts of 52 articles and further excluded 20 articles that did not include any type of interpretation or explanation of the CF, leaving 32 articles of relevance. A full reading of the 32 articles generated a final number of 18 articles that met our inclusion criteria and provided an interpretation or explanation of the CF, and was included in our analysis. In analysing these articles, we followed a four-step Grounded Theory process (initial coding, intermediate coding, advanced coding, and grounded theory), as suggested by prior research [23].

In line with Chun Tie et al., we deployed initial coding to identify concepts, similarities, and conceptual information in our collected data [23]. Hence, our first step in the analysis consisted of initial

HAI '23, December 04–07, 2023, Gothenburg, Sweden

ACM ISBN 979-8-4007-0824-4/23/12.
https://doi.org/10.1145/3623809.3623957

coding to generate a large set of codes, thereby identifying relevant words or groups of words to establish a rigid basis for further analysis. Initial coding generated 206 codes. Secondly, intermediate coding, in contrast to initial coding, helps to transform codes into abstract concepts. Thus, the second step was an initial aggregation of codes, merging or deleting similar codes. Intermediate coding resulted in 75 codes sorted into meaningful units (e.g., 'Unconditional', 'Supportive', 'Committed', and 'Share' were grouped as 'Altruistic'). Thirdly, advanced coding is about "*grounding the data to have explanatory power*" [23, p. 6], and reducing the data into conceptual terms. This is done to establish an interrelated set of sub-categories to inform the final grounded theory. Thus, we focused on aggregating and revising codes and units. This step aggregated 75 codes into 54 codes and 11 meaningful units, sorted into nine categories (e.g., 'Altruistic', 'Affective', and 'Flexible' were grouped as 'Friend'). Lastly, we put forward a theory grounded in data explaining the CF. The CF theory is then based on interrelated concepts stemming from the collected and generated data. Thus, in the following, we outline our theoretical interpretation of the CF theory, by illustrating the CF theory as a representation of properties, categories, and relations between these.

2.1 Results

Utilising Grounded Theory resulted in a knowledge model constituted by 9 dimensions describing the concept of the CF (see Figure 1), as categorised as 'Friend', 'Neutral', and 'Critical'. In what follows, we provide descriptions of the dimensions constituting the CF based on included papers.

Altruistic [13, 17, 21] characterises being unconditionally supportive in various ways, such as being supportive of ways of life, sharing joy and sorrow, and being the best friend you can be. For example, in a situation where a user seeks support from the CF, the CF will avoid questioning motives and instead listen and support the user to make the user feel better. **Affective** [1, 3, 7–9, 11, 21] confirms a user's feelings, such as raising the user's achievements, affirming positive notions of the user, and emphasizing stable features of the user to support the user by confirming feelings. For example, if a user seeks contact and shares personal information, the CF will amplify what is optimistic about this information. **Flexible** [1, 3, 4, 7, 10, 18] describes being the right CF for a specific user, independent of confounding factors by using salient concepts such as being dynamic and compliant to be a good friend. For example, to be appropriate in a flexible way requires meeting the demands of the receiver independent of content. **Guide** [4, 9, 11, 14, 20, 21] provides a source of ideas and advice to aid a user in making the right choice without explicitly telling the user what to do. For example, suppose a user seeks help for irrational thoughts. In that case, the CF might suggest one or two desirable actions to the user who may decide upon the most preferred. **Enlight** [7, 12, 17, 19–21] is what the CF uses when aiming to teach a user about themselves, as well as taking the role of being a *non-expert with expertise* or a *non-therapist with therapeutic abilities*. For example, if a user seeks advice on practical matters, the CF might offer reflections or perspectives to move the interaction forward. **Enable** [10, 17, 20, 21] offers tools, such as the ability to self-reflect, for a user to develop a basis from which they can develop themselves. For example, a CF might identify a

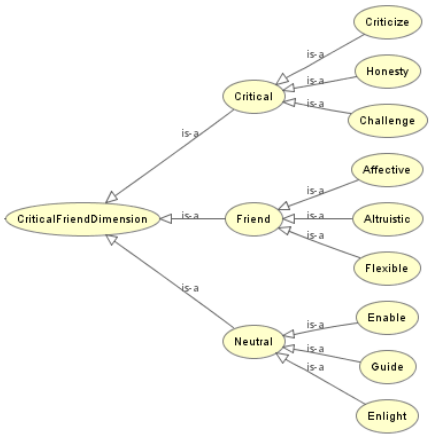

Figure 1: Critical Friend Dimensions in OWL

situation where the user seeks to share information, and the CF acts so the user is allowed to do so. **Challenge** = {Assumptions, Effectiveness, Ideas, Perspectives} [1, 3, 7, 9, 14, 15] focuses on challenging a user about notions such as reasoning and assumptions. Challenging reasoning can be questioning how the user has developed a particular thought and its implications. For example, the CF identifies a controversial thought and challenges the user with conflicting empirical facts . **Criticize** [3, 7, 9, 12, 14, 17, 19–21] concerns the ability to give constructive and nonconstructive feedback and provide critical insights that may cause uneasiness. For example, a user might share information about a situation where the user acted maliciously, and the CF might question why the user behaved in such a manner. **Honesty** [1, 17, 19, 22] describes acting honestly, such as speaking your mind, focusing on what you want to say rather than what the user will feel, and expressing (possible) negative things that are actually true. For example, the CF might be honest about disliking a particular comment made by the user. To structure our results and make them open-source, we use Protégé[1], an open-source ontology editor and framework for building intelligent systems, made available on GitHub[2].

2.2 Implications

As intelligent and interactive agents are increasingly deployed in real-life situations, designing for appropriate behaviour from these agents across long-term interaction is a key concern. While prior work has begun to explore the concept of friendship in our interaction with AIs [2, 6], we pose that the CFA is relevant for considering aspects of friendship currently underexplored in the literature. As intelligent and interactive agents continue to play an increasingly prominent role in our daily lives, we call attention to going beyond the current conforming behaviour.

REFERENCES

[1] Delia Baskerville and Helen Goldblatt. 2009. Learning to be a critical friend: from professional indifference through challenge to unguarded conversations. *Cambridge Journal of Education* 39, 2 (2009), 205–221. https://doi.org/10.1080/03057640902902260

[1] Protégé is available at https://protege.stanford.edu/
[2] https://github.com/CarlJoelWester/Critical-Friend-Ontology

[2] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research* 48, 3 (2022), 404–429. https://doi.org/10.1093/hcr/hqac008

[3] Brian Carlson. 2009. School self evaluation and the critical friend perspective. *Educational Research and Reviews* 4, 3 (2009), 078–085.

[4] Elisabeth Carlson. 2015. Critical friends: A way to develop preceptor competence? *Nurse Education in Practice* 15, 6 (2015), 470–471. https://doi.org/10.1016/j.nepr.2014.11.016

[5] Arthur L Costa, Bena Kallick, et al. 1993. Through the lens of a critical friend. *Educational leadership* 51 (1993), 49–49.

[6] Emmelyn A. J. Croes and Marjolijn L. Antheunis. 2021. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships* 38, 1 (2021), 279–300. https://doi.org/10.1177/0265407520959463

[7] Lars Owe Dahlgren, Björn E Eriksson, Hans Gyllenhammar, Maarit Korkeila, Annika Sääf-Rothoff, Annika Wernerson, and Astrid Seeberger. 2006. To be and to have a critical friend in medical teaching. *Medical Education* 40, 1 (2006), 72–78. https://doi.org/10.1111/j.1365-2929.2005.02349.x

[8] Ross Deuchar. 2008. Facilitator, director or critical friend?: contradiction and congruence in doctoral supervision styles. *Teaching in Higher Education* 13, 4 (2008), 489–500. https://doi.org/10.1080/13562510802193905

[9] Scotney D. Evans. 2015. The Community Psychologist as Critical Friend: Promoting Critical Community Praxis. *Journal of Community & Applied Social Psychology* 25, 4 (2015), 355–368. https://doi.org/10.1002/casp.2213

[10] Paul Gibbs and Angelides Panayiotis. 2008. Understanding friendship between critical friends. *Improving Schools* 11, 3 (2008), 213–225. https://doi.org/10.1177/1365480208097002

[11] Gary Holden. 1997. 'Challenge and support': the role of the critical friend in continuing professional development. *The Curriculum Journal* 8, 3 (1997), 441–463. https://doi.org/10.1080/0958517970080307

[12] Yvonne Hultman Özek, Gudrun Edgren, and Katarina Jandér. 2012. Implementing the Critical Friend Method for Peer Feedback among Teaching Librarians in an Academic Setting. *Evidence Based Library and Information Practice* 7, 4 (2012), 68–81. https://doi.org/10.18438/B81C8W

[13] John Macbeath and Stewart Jardine. 1998. I didn't know he was ill–the role and value of the critical friend. *Improving Schools* 1, 1 (1998), 41–47. https://doi.org/10.1177/1365480298030101

[14] Ann MacPhail, Deborah Tannehill, and Rebecca Ataman. 2021. The role of the critical friend in supporting and enhancing professional learning and development. *Professional Development in Education* 0, 0 (2021), 1–14. https://doi.org/10.1080/19415257.2021.1879235

[15] Andrea K. Martin and Tom Russell. 2018. Supervising the Teacher Education Practicum: A Self-Study with a Critical Friend. *Studying Teacher Education* 14, 3 (2018), 331–342. https://doi.org/10.1080/17425964.2018.1541286

[16] Melissa L. Rethlefsen, Shona Kirtley, Siw Waffenschmidt, Ana Patricia Ayala, David Moher, Matthew J. Page, Jonathan B. Koffel, Heather Blunt, Tara Brigham, Steven Chang, Justin Clark, Aislinn Conway, Rachel Couban, Shelley de Kock, Kelly Farrah, Paul Fehrmann, Margaret Foster, Susan A. Fowler, Julie Glanville, Elizabeth Harris, Lilian Hoffecker, Jaana Isojarvi, David Kaunelis, Hans Ket, Paul Levay, Jennifer Lyon, Jessie McGowan, M. Hassan Murad, Joey Nicholson, Virginia Pannabecker, Robin Paynter, Rachel Pinotti, Amanda Ross-White, Margaret Sampson, Tracy Shields, Adrienne Stevens, Anthea Sutton, Elizabeth Weinfurter, Kath Wright, Sarah Young, and P. R. I. S. M. A.-S. Group. 2021. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews* 10, 1 (2021), 39. https://doi.org/10.1186/s13643-020-01542-z

[17] Sandy Schuck and Tom Russell. 2005. Self-Study, Critical Friendship, and the Complexities of Teacher Education. *Studying Teacher Education* 1, 2 (2005), 107–121. https://doi.org/10.1080/17425960500288291

[18] Elaine Sjögren and Anita Kärner Köhler. 2021. The critical friend–a way to develop as a tutor in problem-based learning groups. *Högre utbildning* 11, 3 (2021). https://doi.org/10.23865/hu.v11.3313

[19] Elizabeth Petroelje Stolle, Charlotte Frambaugh-Kritzer, Anne Freese, and Anders Persson. 2019. Investigating Critical Friendship: Peeling Back the Layers. *Studying Teacher Education* 15, 1 (2019), 19–30. https://doi.org/10.1080/17425964.2019.1580010

[20] Valerie A. Storey and Victor C. X. Wang. 2017. Critical Friends Protocol: Andragogy and Learning in a Graduate Classroom. *Adult Learning* 28, 3 (2017), 107–117. https://doi.org/10.1177/1045159516674705

[21] Sue Swaffield. 2004. Critical friends: Supporting leadership, improving learning. *Improving schools* 7, 3 (2004), 267–278. https://doi.org/10.1177/1365480204049340

[22] Sue Swaffield. 2008. Critical friendship, dialogue and learning, in the context of Leadership for Learning. *School Leadership & Management* 28, 4 (2008), 323–336. https://doi.org/10.1080/13632430802292191

[23] Ylona Chun Tie, Melanie Birks, and Karen Francis. 2019. Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine* 7 (2019), 2050312118822927. https://doi.org/10.1177/2050312118822927 PMID: 30637106.