The Artificial Intelligence (AI) Hermeneutic Network:
A New Approach to Analysis and Design of Intentional Systems

Jichen Zhu and D. Fox Harrell, Ph.D.
School of Literature, Culture, and Communication | Georgia Institute of Technology
jichen.zhu@lcc.gatech.edu, fox.harrell@lcc.gatech.edu

"I felt that I should be able to get the computer to sound good more or less on its own, so that someone listening to it says, 'Who is that playing?' But if you get `What's that?' instead, you have to go back to the drawing board."

(Lewis, 2000)

Abstract

Digital information technologies are increasingly being adopted in the humanities as both research tools and supports for new forms of cultural expression. Some of these digital technologies, in particular artificial intelligence (AI) programs, exhibit complex behaviors described as creativity, planning, and learning, behaviors usually seen as the territory of intentional human phenomena. This paper identifies a prototypical subset of these programs, which we name *intentional systems*, and argues that their seemingly "intentional" behaviors are not the sole effect of underlying algorithmic complexity and knowledge engineering practices from computer science. In contrast, we argue (paralleling the field of software studies) that intentional systems, and digital systems at large, need to be analyzed as a contemporary form of historically, culturally, socially, and technically situated *texts*. Perception of system intentionality arises from a network of continuous meaning exchange between system authors' narration and users' interpretation processes, situated in social contexts. The central contribution of this paper is a new interdisciplinary analytical framework called the *AI hermeneutic network* that is informed by traditions of hermeneutic analysis, actor-network theory from sociology and science studies, cognitive semantics from cognitive science, computational semantics from

computer science, and philosophy of mind. To illustrate the functioning of the AI hermeneutic network we present our recent work called the *Memory, Reverie Machine*, an expressive intentional system that generates interactive narratives rich with daydreaming sequences.

Intentional Systems

Trombonist and composer George Lewis's above description of his interactive musical system *Voyager* exemplifies a growing number of digital systems, such as the autonomous painting program *AARON* (Cohen, 2002) and recent computational narrative works (Mateas and Stern, 2002, Pérez y Pérez and Aliseda, 2006, Harrell, 2006) that utilize AI techniques in pursuit of cultural expression. Decades after heated debates about the feasibility of AI, the question of whether computers may one day possess human-level intelligence need no longer spur society's fear and curiosity. Instead, systems are designed to encourage users to make sense of them as intentional and independent entities. Compared to instrumental, production-oriented, programs such as the *PhotoShop*, these systems display intentional behaviors related to human mental phenomena such as "planning," "learning," "narrating," and "creating," as if their actions were *about* something in the world (Searle, 1983) rather than mere execution of algorithmic rules. Lewis, for instance, insists that *Voyager* "not [be] treated as a musical instrument, but as an independent improviser." He deliberately designed the system to display independent behaviors arising from its own internal processes that even its designer cannot fully anticipate. The improvisational *dialogue* between *Voyager* and the musicians, Lewis emphasizes, is "bi-directional transfer of intentionality through sound."

This paper proposes a definition of *intentional systems* as those that incorporate: 1) significant computational complexity, 2) process opacity, 3) human-like coherent behaviors, and 4) execution of authorial intention. The term encompasses not only AI systems but also AI-like systems that exist

either outside of computer science communities or are not described by their authors as AI systems for ideological or other reasons. Critical analysis and design of intentional systems, like information technologies at large in the digital humanities, calls for the recognition of these systems as important forms of cultural production, beyond their traditionally instrumentalized, productivity oriented roles.

Intentional Systems as Texts

Although generally used to describe written forms of discourse, the term *text* as the object of literary theory and modern hermeneutics is not confined to only linguistic forms. In his essay on the literary text, German philosopher Manfred Frank (Frank, 1989) criticizes the notion that meanings that authors encode within texts can be objectively and without distortion be retrieved by readers given an appropriate methods of interpretation (Hirsch, 1967). Instead, Frank proposes a complex communication process in which both author and reader actively create, shape, and reconstruct meanings. This echoes the even broader notion of dialogic meaning posited by the Russian philosopher and critic Mikhail Bakhtin in which language is understood as dynamic, contextual, intertextual, and relational (Holquist, 1990). Acknowledging the textuality of intentional systems opens up understanding of system intentionality to a range of socially situated methods.

Intentional systems are not simply the result of clever algorithmic and data structural innovations. The AI practitioner and theorist Philip Agre cogently points out that the "the purpose of AI is to build computer systems whose operation can be narrated using intentional vocabulary." (Agre, 1997) Michael Mateas, co-developer of *Façade,* further deconstructs the codes invoked in AI practice by raising a distinction between the "rhetoric machine" (e.g. discursive strategies, humanly interpreted computation, and definitions of system progress) and co-existing "code machine" (including physical processes, computational processes, and complex causal flow), in order to pin down the

long-neglected social and discursive aspect of AI systems. (Mateas, 2002) In addition to considering actual computer programs, analysis of intentional systems should not omit the authors' publications, presentations, and interpersonal communication about the system. Such narrative output is historically and culturally situated and must be considered as part of the intentional system.

The AI Hermeneutic Network

The central contribution of this paper is an interdisciplinary framework called the AI hermeneutic network, enabled by theorizing intentional systems as texts, to analyze system intentionality as a result of a hermeneutic communication process that involves both authors' narrations and users' interpretations through interaction with both actual systems and authors' written and verbal output. In addition, this paper recognizes that intentional systems exist in social contexts that involve more than just authors and users. Animate and inanimate actors, called "actants" in actor-network theory (Latour, 1996, Callon, 1986), participate in the network through multi-directional communication. Government and military funding, for instance, often plays a prominent role in determining direction and validity of different approaches of AI research.

Historically, hermeneutic studies developed interpretative theories and methods in order to recover the meanings of sacred texts intended by the (divine) author(s). Modern hermeneutics, influenced by Schleiermacher, recognizes that *everything* calls for the work of interpretation and broadens itself to the philosophical interrogation of interpretation. (Virkler, 1981) This paper highlights discursive strategies in authors' narrations of intentional systems. For instance, Agre calls attention to the "elasticity" of the AI key words, such as "planning". (Agre, 1997) He observes that these key terminologies are simultaneously precise (formal) and vague (vernacular), which allows AI

practitioners to seamlessly integrate their everyday experience as embodied intentional being in the algorithmic research, and to narrate computation with popularly accessible vernacular vocabulary.

One relatively unexplored aspect of this continuous chain of translation of values and meanings between both human and computational actors (Latour, 1996) is users' readings and interpretations of intentionality from systems that are clearly inanimate. For example, the interpretations of *Voyager*'s behaviors as intentional formed by its human co-performers and their audiences are central to construe *Voyager's* intended status as a performer in its own right, as intended by Lewis. As Frank argues that "[i]n the understanding of its readers the text … acquires a meaning which exceeds the memory of its origin,"(Frank, 1989, p.39) the analysis of system intentionality is not adequate without considering participation of users and audiences.

This paper emphasizes the discursive strategy and semantic interpretation from a cognitive linguistics perspective, as opposed to arising as observable phenomenona of AI algorithms and knowledge structures alone. Conceptual blending theory (Turner, 1996) (Fauconnier, 2001) (Fauconnier and Turner, 2002) offers a cognitive foundation for understanding system intentionality as actively (re)constructed by users via integrating concepts of intentionality encountered with animate objects, with conceptualization of algorithmic operation of computer systems. Thus, the various discourses surrounding intentional systems provide semantic hooks with which users are able to compress the behavior of unfamiliar computational systems to human scale by constructing blends of intentional systems and more familiar animate agents.

Conclusion: Design Implications of the AI Hermeneutic Network

The novel framework of the hermeneutic network suggests new design approaches for intentional systems in digital humanities. Our current interactive narrative work *Memory, Reverie Machine*

generates stories in which the main character varies dynamically along a scale between a user-controlled avatar with low intentionality and an autonomous non-player character with high intentionality. By algorithmically controlling the semantic hooks for interpreting system behavior s intentional in the discursive structure (Zhu and Harrell, 2008), the system turns system intentionality into a scalable expressive tool in computational narrative.

In conclusion, this paper proposes a new interdisciplinary framework to analyze intentional systems as social and cultural productions, as opposed to construing them as the domain of purely technical practices. It underlines authors' narrative and users' interpretative strategies, in a socially situated network of meaning exchange. Finally, through our own computational work we suggest new design implications for intentional systems, such as the scale of intentionality (Zhu and Harrell, 2008) that potentially can add new forms of expressivity to creative digital discourse systems.

References

AGRE, P. E. (1997) Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. IN BOWKER, G. C., STAR, S. L., TURNER, W., GASSER, L. & BOWKER, G. (Eds.) *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide.* Lawrence Erlbaum Associates.

CALLON, M. (1986) Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. IN LAW, J. (Ed.) *Power, Action and Belief: A New Sociology of Knowledge.* London, Routledge & Kegan Paul.

COHEN, H. (2002) A Self-Defining Game for One Player: On the Nature of Creativity and the Possibility of Creative Computer Programs. *Leonardo,* 35**,** 59-64.

FAUCONNIER, G. (2001) Conceptual Blending and Analogy. IN GENTNER, D., HOLYOAK, K. J. & KOKINO, B. N. (Eds.) *The Analogical Mind: Perspectives from Cognitive Science.* Cambridge, MA, MIT Press.

FAUCONNIER, G. & TURNER, M. (2002) *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*, Basic Books.

FRANK, M. (1989) *The Subject and the Text: Essays on literary theory and philosophy,* Cambridge, Cambridge University Press.

HARRELL, D. F. (2006) Walking Blues Changes Undersea: Imaginative Narrative in Interactive Poetry Generation with the GRIOT System. *AAAI 2006 Workshop in Computational Aesthetics: Artificial Intelligence Approaches to Happiness and Beauty.* Boston, MA, AAAI Press.

HIRSCH, E. D. (1967) *Validity in Interpretation* New Haven and London, Yale University Press.

HOLQUIST, M. (1990) *Dialogism: Bakhtin and His World.*

LATOUR, B. (1996) *Aramis, or the Love of Technology,* Cambridge, Harvard University Press.

MATEAS, M. (2002) Interactive Drama, Art, and Artificial Intelligence. *Computer Science.* Pittsburgh PA, Carnegie Mellon University.

MATEAS, M. & STERN, A. (2002) A Behavior Language for Story-Based Believable Agents. *IEEE Intelligent Systems,* 17**,** 39-47.

PÉREZ Y PÉREZ, R. & ALISEDA, A. (2006) The Role of Abduction in Automatic Storytelling. *Proceedings of the AAAI workshop in Computational Aesthetics: AI Approaches to Beauty & Happiness.* Boston, MA, AAAI Press.

SEARLE, J. (1983) *Intentionality: An Essay in the Philosophy of Mind,* Cambridge, Cambridge University Press.

TURNER, M. (1996) *The Literary Mind: The Origins of Thought and Language,* New York; Oxford, Oxford UP.

VIRKLER, H. A. (1981) *Hermeneutics: Principles and Processes of Biblical Interpretation,* Grand Rapids, MI, Baker Book House.

ZHU, J. & HARRELL, D. F. (2008) Daydreaming with Intention: Scalable Blending-Based Imagining and Agency in Generative Interactive Narrative. *AAAI 2008 Spring Symposium on Creative Intelligent Systems.* Stanford, CA, AAAI Press.