Michael Mateas

Research Statement

AI-based Art and Entertainment

My work is in Artificial Intelligence (AI)-based art and entertainment. I simultaneously engage in AI research and art making, a research agenda and art practice I call *expressive AI* (Mateas 2001).

Expressive AI has two major, interrelated thrusts: (1) *exploring the expressive possibilities of AI architectures* - posing and answering AI research questions that wouldn't be raised unless doing AI research in the context of art practice, and (2) *pushing the boundaries of the conceivable and possible in art* - creating artwork that would be impossible to conceive of or build unless making art in the context of an AI research practice.

1. ***Office Plant #1***

Office Plant #1 (collaboration with Marc Boehlen) is a desktop robot which physically responds with ambient movements and sounds to the social and emotional tone of email received by its user (Boehlen and Mateas 1998). *Office Plant* makes use of a fuzzy cognitive map for behavior selection and statistical methods for email processing.

2. Title screen of *Terminal Time*

Terminal Time (collaboration with Paul Vanouse and Steffi Domike) is an interactive theater system which constructs ideologically biased documentary histories in response to audience feedback measured by an applause meter (Mateas, Vanouse and Domike 2000). *Terminal Time* makes use of formal reasoning techniques to generate the biased histories.

Expressive AI: AI Agenda

The field of Artificial Intelligence is the modern incarnation of an age old quest or dream, the dream of building an image of the human in the machine. It is this dream, fueled by science fiction representations of AI such as Hal 9000 or Commander Data, which is the initial inspiration for many researchers entering the field. This dream is not just about modeling rational problem solvers, but about building machines which in some sense engage us socially, have emotions and desires, and, through our interactions with them, tell us something about ourselves. AI is a way of exploring what it means to be human by *building systems*. An AI architecture is a machine to think with, a concrete theory and representation of some aspect of the human world. Art also explores what it means to be human by *building concrete representations* of some aspect of the human world. Artists often explore aspects of humanity which have been under-explored or ignored in AI research. What Joseph Bates wrote about character animators in particular, applies to the arts in general:

> "It can be argued that while scientists may have more effectively created [machines which act like] scientists, it is the artists who have come closest to understanding and perhaps capturing the essence of humanity that … AI researchers ultimately seek." (Bates 1994)

Combining these two ways of knowing-by-making opens a new path towards the AI dream, a path which takes seriously the problem of building intelligences that robustly function outside of the lab to engage human participants in intellectually and aesthetically satisfying interactions which, hopefully, teach us something about ourselves.

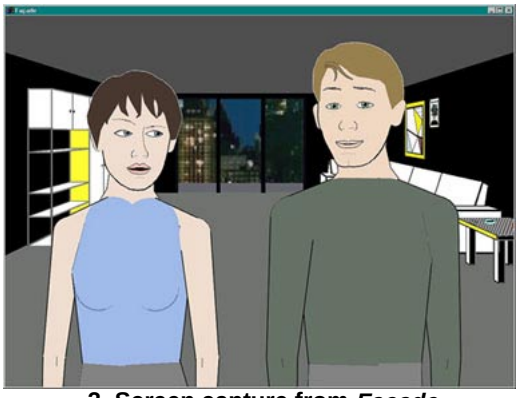
Rather than focusing on isolated competencies such as planning, classification, or logical inference, **I build complete architectures, that is, AI systems which integrate representations, processes and perception to engage in expressive behavior**. By "expressive behavior" I mean behavior which can be read by an audience and thus allows the audience to interact meaningfully with the system. **This focus on building complete, expressive architectures opens a number of research areas including believability, narrative intelligence, perception of meaningful situations, manipulation of non-language-like representations such as music and images, and models of creativity.** Each of these areas clusters a number of specific research questions. Below I briefly describe three of these areas, believability, narrative intelligence, and perception of meaningful situations, in the context of four AI-based artworks.

Believability

When Turing introduced his famous test for intelligence, he also introduced a subversive and not always recognized idea: *that intelligence is not a property of a system itself, but rather resides in the details of an interaction and the perceptions of an observer*. The idea of believability, which was introduced into AI discourse by the Oz project (Bates 1994), is, like the Turing Test, an observer-centric notion. However, instead of focusing on imitating the responses of a "generic human," believability focuses on the idea of character, on rich and compelling presentations of behavior which foster the willing suspension of disbelief. Where the Turing Test is about closing the gap between the real and not-real (building systems which are indistinguishable from a real human), *believability is about building autonomous agents which function as-if-real*, in the same way that characters such as Hamlet or the Terminator can't be described unequivocally as real or fake, but rather function as-if-real within their respective worlds. **In expressive AI, I generalize the notion of believability beyond human-like characters to the agenda of building AI systems that engage in internally-consistent, evocative and compelling behaviors** which encourage participants to suspend disbelief and interact with the system.

Believability research questions include:

- For believable behavior, what fit must exist between the physical or virtual representation of a system and the behavior architecture?
- What kinds of architectures support human-like characters?
- What sorts of procedural and declarative representation frameworks are appropriate for supporting the human authoring of believable behavior?



3. Screen capture from *Façade*

Façade (collaboration with Andrew Stern) is an interactive drama combining both autonomous, believable characters with dynamic plot construction (Mateas and Stern 2002). The player has been invited over to a couple's apartment for drinks, where she ends up participating in their marital troubles.

Façade (figure 3) explores the integration of character and story within a dramatic world. Most work in believable agents has been organized around the metaphor of strong autonomy. Such agents choose their next actions based on local perception of the environment plus internal state corresponding to the goals and possibly the emotional state of the agent. However, for believable agents participating in interactive stories, strong autonomy is problematic; a character's behavior is not only organized around private goals and local state but also around story goals and global story state. Yet centralized deliberative control of *all* behavior would not support the moment-by-moment decisions characters must make. *Façade* resolves this problem through weakly autonomous characters which are tightly coupled to a more deliberative story-guidance framework while still maintaining the reactivity necessary for real-time interaction.

Office Plant #1 (figure 1) provides a physical, sculptural manifestation of email activity, inviting its owner to contemplate the ebb and flow of her communications throughout the day. To physically manifest this email flow requires a sensitivity to the local email history as well as individual emails. This requirement, along with the fact that a

single email may belong to multiple social and emotional categories and multiple emails may be simultaneously received, means that simple, stateless mappings from email categories to behavior are inappropriate. Additionally, the ambient nature of *OP#1*'s movements means that the readable behavior of the plant's body are its "poses," while the movement between poses should be very slow and subtle. A fuzzy cognitive map (similar to a recurrent neural net), in which each node corresponds to a pose, satisfies these requirements, maintaining local history, resolving conflicting behavior tendencies through winner-take-all competition, and producing long-term behavior consisting of trajectories of poses.

Narrative Intelligence

Story-telling and narrative are fundamental to human experience. As children we are immersed in stories and learn to approach the world via narrative frameworks. As adults we order events and find meaning by assimilating them to more-or-less familiar narratives. Given the fundamental role narrative plays in both human intelligence and art (Mateas and Sengers 2002), my work necessarily involves the construction of narrative systems.

Narrative intelligence research questions include:

- What representations and architectures are appropriate for story generation and story guidance in interactive story systems?
- How can intelligent systems help us to more deeply understand the stories we tell?
- What role can narrative play in architectures for autonomous agents?

Terminal Time (figure 2) is a story generation system which constructs ideologically-biased documentary histories, consisting of spoken narrative, video sequence and sound track, in response to audience feedback. One of the goals of *Terminal Time* is to build a caricature model of the documentary film production process. Rather than "objectively" reporting a sequence of events through the eye of a camera (the implied production process in documentary film), events are instead selected and biased so as to satisfy an ideological position, assembled into a desired narrative, and only then is video footage selected to illustrate the constructed narrative. In order to model this process the architecture makes use of several representations and knowledge sources including: a knowledge base of historical events, represented in an ontology based on the Upper Cyc Ontology, ideologue-specific representations of rhetorical goals which select and "spin" events, rhetorical devices which can be used to "glue" spins together to form historical narratives, a plan-based natural language generator, and a database of term-indexed video clips.

4. Diagram of *Dinner Table*

The planned future project *Dinner Table* (collaboration with Marc Boehlen) is a robotic board game embedded in a table. Game-like narratives are played out on the surface of the table, using a combination of autonomous physical piece movement and photographic imagery, in response to real-time features of the conversation at the table.

Façade (figure 3) is an interactive drama which incorporates the player's interaction with autonomous characters into a well-shaped dramatic arc with a clear inciting incident, progressive complication leading to a climax, and closure. The dramatic guidance architecture must represent the story and interpret interaction using "story pieces" which are small enough to support interesting combinatorial possibilities while respecting the constraints of dramatic arc construction. The theory of dramatic writing suggests a candidate "story piece" in the dramatic beat. In dramatic writing, a beat is the smallest unit of dramatic value change, where dramatic values are properties of individuals or relationships such as trust, love, hope, etc. In *Façade* beats are architectural entities, consisting of preconditions, a description of the values changed by the beat, success and failure conditions, and joint behaviors which coordinate the characters in order to carry out the specific beat. The drama manager attempts to sequence beats so as to incorporate player interaction while making specific dramatic arcs (value change graphs) happen.

Perception of Meaningful Situations

Humans operate in contexts rich with language, actions and events, all of which are laden with meaning. **In order to participate in human contexts, an AI-based artwork must be able to perceive human meanings**.

Perception of meaning research questions include:

- How must machine learning techniques be adapted for less well-defined classification tasks?
- What features of the environment can serve as proxies for contextual meanings which are not amenable to direct sensing?
- What architectural balance must be achieved so that there is a tight "fit" between what can be sensed and the expressive behavior that can be generated?

Office Plant #1 (figure 1) uses classification techniques such as naïve Bayes and k-nearest-neighbor to sort incoming email into social and emotional categories such as `thank you`, `request`, `apology`, and `chatty`. By performing such classifications, *OP#1* is able to perceive part of the office context in which it is placed. Text classification techniques are typically applied to content classifications (e.g. is this a CS or biology article) rather than "tone" classification (e.g. is this email chatty or intimate). *OP#1* pushes text classification techniques by attempting multi-label "tone" classifications.

The future piece ***Dinner Table*** (figure 4) responds to the conversational dynamics of two dining companions at the table. The challenge is to identify potentially interesting auditory and visual features of dinner table conversation and figure out how to detect them. Candidate features include laughter, the arrival of food, the ratio of time one person spends talking versus another, and periods of active eating.

Expressive AI: Art Practice

Artists can engage computer technology in a number of ways. One way is to use the computer as a tool to manipulate and generate images, sounds, and interactive multimedia presentations. This does not require a deep engagement with computer science, but rather a facility with tools written by others. Another mode of engagement is work which is *about* science and technology. Such work appropriates well understood techniques, ideas and tools to make malfunctioning assemblages which (often humorously) expose the ideological underpinnings of the technology. However, such work does not move forward in offering new, positive dimensions for techno-scientific development. **As an artist, I employ a third mode of engagement with AI in which art practice is seen as research, as a mode of inquiry generating new knowledge within AI**. I view AI as an expressive medium. As an active researcher in the field, I am able to explore the potential of this medium with an acute sensitivity to the complex patterns of resistance and opportunity in the material realization of intelligence. Here I briefly explore three practical and theoretical issues which arise in expressive AI as an art practice. For more on expressive AI as an art practice see (Mateas 2001).

Necessity of concrete grounding

Expressive AI must be done in the concrete context of building specific AI-based artworks. Research questions and technical innovations co-evolve within a specific conceptual and aesthetic context. For example, *Office Plant #1* was originally conceived as a robotic office companion, a machine offering a commentary and opportunity for reflection on the office scene. Architectural explorations of various sensing and behavior selection mechanisms co-evolved with changes and refinements of the concept and physical instantiation, resulting in a final "meshing" of concept and architecture in the form of an email-sensitive ambient sculptural presence reminiscent of a plant form.

Architectural Affordances

In an AI-based artwork, the AI architecture itself is a fundamental part of the piece. **The architecture determines the relationship between the artist and the audience by determining the relationship between interpretive and authorial affordances**. The interpretive affordances are the "hooks" the system makes available to an audience for making sense of the piece, for constructing meaning, and for understanding interaction with the piece. The authorial affordances are

the “hooks” the system makes available to the artist for inscribing their artistic intention on the machine. The architecture determines how the bits of pieces of authored content (e.g. behaviors, inference rules, video clips, etc.) come together to produce audience-interpretable activity. Architectural design decisions, such as the representation of beats in *Façade* or rhetorical devices in *Terminal Time*, are thus fundamentally intertwined with the concept of the piece.

In much electronic media practice the internal structure of the machine is generally marginalized. The machine itself is considered a hack, an accidental byproduct of the artist's engagement with the concept of the piece. Why would an artist want to concern herself with authorial affordances, with the structural properties of the machine itself? Because such a concern allows an artist to explore expressive possibilities that can only be opened by a simultaneous inquiry into interpretive affordance and the structural possibilities of the machine. An artist engaging in expressive AI practice will be able to build computer-based works with a depth, a richness, and a sophistication that can't be achieved without this simultaneous focus on meaning-making and machine structure.

Video Games as an Art Form

AI-based interactive art has the potential to hybridize with computer games to form a new mass-audience high-art form. Already in game marketing “game AI” is becoming an increasingly important differentiator, with breakthrough games such as *The Sims* and *Black and White* being noted primarily for their AI. The emerging academic field of games studies points to the increasing cultural relevance of games. **I contribute to the growing body of work in game design and games studies, helping to create a new AI-based art form**.

Conclusion

Expressive AI is a radically interdisciplinary endeavor, combining deep study of AI architectures with art practice and humanistic studies of culture. **By combining two ways of knowing-by-building, expressive AI simultaneously opens a new path in the AI quest of understanding ourselves by building systems, and opens up new realms of human expression and experience**. Rather than singling out a particular technical tradition, expressive AI is a stance or viewpoint from which all of AI can be rethought and transformed. I’m interested in building a research group which would bring together interdisciplinary elements to create an energetic zone pushing both AI research and art in vital, new directions.

References

Bates, Joseph. 1994. The Role of Emotion in Believable Agents. *Communications of the ACM, Special Issue on Agents*, July 1994.

Boehlen, M. and Mateas, M. 1998. Office Plant #1: Intimate space and contemplative entertainment**.** *Leonardo: Journal of the International Society for Arts, Sciences, and Technology*, 31 (5), 1998. 345-348.

Mateas, M. and Senger, P. (Eds.). 2002 (forthcoming). Narrative Intelligence. Amsterdam: John Benjamins.

Mateas, M. and Stern A. 2002 (forthcoming). Towards Integrating Plot and Character for Interactive Drama**.** In K. Dautenhahn (Ed.), *Socially Intelligent Agents: The Human in the Loop*. Kluwer.

Mateas, M. 2001. Expressive AI: A hybrid art and science practice**.** *Leonardo: Journal of the International Society for Arts, Sciences, and Technology* 34 (2), 2001. 147-153.

Mateas, M., Vanouse, P., and Domike S. 2000. Generation of Ideologically-Biased Historical Documentaries**.** In *Proceedings of AAAI 2000*, 36-42. Austin, TX.