AI and Cinema - Does artificial insanity rule?

Robert B. Fisher
Division of Informatics
University of Edinburgh

Robots and other artificially intelligent (AI) mechanisms are now a common plot device or even main character in movies [1, 2, 5, 6, 7, 8]. What is less well known is that they have appeared since the early days of cinema (1907). These movies are interesting, because they help shape the mainstream public's view of artificial intelligence and robotics. The experienced science fiction reader and AI professional have more developed views about AI, but these are minority of the population. Here I'm interested in how AI is understood and interpreted by the majority.

Continuing scientific developments are already bringing AI research results to the public: web search engines, network routing, automatic scheduling programs, automobile route planning, speech understanding, factory automation, home robot vacuum cleaners and lawn mowers, etc. But these seem like normal products and are not usually viewed as Artificial Intelligence, at least as AI appears in the cinema. The cinematic sort of AI is the key opinion former of AI: a challenge to humans because of their potential powers. This essay considers why these powerful cinematic AI agents, in spite of their presumed abilities, have largely abnormal "personalities".

A recent www-based survey of 256 movies containing elements of AI [4] has 83 "true" AI entries. From the 256 movies, I have excluded TV episodes (too numerous), short cartoons (too numerous), cyborgs (36), androids (20) and still unclassified movies (117). A debate could be held around the excluded cyborgs and androids, however, my concern here is with true Artificial Intelligence: the product of a non-human sensing and reasoning mechanism. Cyborgs have human brains (although perhaps physically or computationally augmented), so do not say anything about AI. Androids are more ambiguous as they are constructed entities. They are generally "grown" to be either identical to humans or "improved" versions. However, their mental machinery is assumed to be identical to humans, and so the use of an android says nothing about the implications of having an artificial reasoning mechanism. I have excluded these cases as also being uninteresting since they say nothing about the nature of non-human thought.

Of the 83 "true AI" movies, 46 depict a "mindless" sort of AI without self-reflection or self-awareness, largely as efficient robotic killing machines. Although this is an exciting plot device and a worrisome possible future, these AI agents do not have much depth and ultimately will interest viewers at about the same level as a well-engineered automobile. (Movies with mindless AI agents are listed in Appendix A.)

But what of the other 37 movies (listed in Appendix B)? Examples of these are Metropolis (1926), The Wizard of Oz (1938), 2001: A Space Odyssey (1968), Demon Seed (1977), Making Mr. Right (1987), Virtuosity (1997) and AI Artificial Intelligence (2001). These movies have AI-based agents as central characters, interacting with humans in many of the same ways as humans. They may not always be a cinematic or commercial success, but, because they explore the nature of AI agents, how the agents minds may work and the consequences of the integration of the AI agents into human society, they provide a glimpse at how the rest of society views AI agents. This view is not a healthy one.

What's Special about AI

A short summary of the some technical aspects of Artificial Intelligence would help to make clear the central issue of this article. If we are considering the sorts of active AI entities typical to science fiction works, whether bodied or electronic, then these agents are likely to have classical AI structures, at least when viewed from an external viewpoint (if not also internally as part of the agent's implementation). These structures include: a perceptual mechanisms, a world knowledge database, a short-term working memory subsystem, common-sense knowledge, a planning ability, an actuator mechanism (*e.g.* a speech or robotic motion subsystem) and a reasoning mechanism.

Central to this article is the reasoning mechanism. If we augmented a human with an infrared vision system or a mechanical prosthesis, then we wouldn't think of the human as being more than physically enhanced. Similarly, if we added access to a larger knowledge base or greater memory, we would be impressed with the human's abilities, but would still consider him/her to be enhanced but essentially human. Thus, if anything, it is the reasoning mechanism that sets the AI agent apart from humans.

The classical understanding of AI reasoning mechanisms is that they are all variants of logical reasoning. These include a variety of pattern matching and manipulation systems (*e.g.* production systems) or explicit logic (*e.g.* of the familiar form "All men are mortal" plus "Socrates is a man" therefore "Socrates is mortal"). Mathematically, all known sophisticated reasoning mechanisms are equivalent (*i.e.* there is nothing that you can conclude using one method that can't be concluded when using another method, although you might be able to do it faster using some methods). The mechanisms have also been proven mathematically to be fundamentally limited (*i.e.* Gödel showed that there are things that can be seen to be true by humans in even simple formal reasoning systems but these things cannot ever be proved). Current computers are based on these logical mechanisms, so some people argue that they can never be intelligent. At least they can never act like humans, since humans never act so perfectly logically.

Much has been made of the recent success of the computer program Deep Blue's challenge to the human world chess champions. But, Deep Blue's skill comes from a very different mechanism than human skills: it is extraordinarily fast, has special computer components specifically for chess playing, has access to thousands of stored previous game situations and has perfect memory of its reasoning. Humans cannot match this, but have

better judgement about what is or is not a likely strategy, and have good heuristics ("rules of thumb") about what to do and when. Deep Blue can only play chess.

Chess and mathematics are domains where "exact" reasoning skills are taken to the limits, but these are exactly where computers have strengths. When it comes to other activities, such as teaching a person a new skill, human reasoning approaches (so far) are much more successful. When playing chess, we can recognize the essential details of a situation and its similarity to previous known situations, know various strategies that might be applicable, and have judgement about what might work. We also make lots of mistakes, which can lead to new discoveries that a more formally correct machine might not make.

So, to give an exaggerated summary of the point: human and AI reasoning mechanisms are currently and are likely to remain quite different. Thus, the consequences of that reasoning will be different - the choice of evidence used, decisions made and actions taken might be quite different for an AI agent, and thus unintelligible to humans.

Wooden Acting

The roles that AI agents hold in the movies are largely the same as those that humans might hold, *e.g.*:

- Comic relief - Star Wars (R2D2 & C3PO)
- Background - Red Planet (the spaceship)
- Antagonist - The Terminator, The Matrix
- Protagonist - Bicentennial Man
- Companion - Knight Rider (the car)
- Assistant - Silent Running (gardening robots)

There is nothing particularly special about how an AI agent might contribute to a movie. In fact, the plots and character roles of movies containing AI agents are hardly different than mainstream movies: *e.g.* action movies structured around a crisis and its resolution, rite of passage movies where the main character extends his/her status, abilities or outlook, extreme behavior movies that explore the interaction between abnormal behavior and society, etc.

Are these movies realistic in their depiction of AI? This depends a lot on what one believes is realistic, but at least several criteria are applicable:

1. The abilities that AI agents have are physically realistic by current standards. This allows different agent sizes, broader and faster communication and increased strength. Unrealistic abilities would include invisibility, mind reading, time travel, etc.

2. The agent's sensing and information gathering processes may be more widespread and distributed, have extended perceptual ranges and sensitivities, and gather more information, but are not omnipresent.

3. The agent's reasoning processes may be faster, more thorough, incorporate more information, but are not omniscient.

The key justification for believability is that the technologies used to construct the AI agents are plausible, but significant extensions of existing technologies. There might be some theoretical limits on the language abilities, personalities, intellectual capabilities of a plausible AI agent, but these seem largely without practical limitation. By these criteria, most AI agents presented in the cinema are close to being realistic. The mindless killing machines and advanced household and driving appliances are easily plausible AI agents, but the AI agents with 'minds' might be contentious. My opinion is there is nothing obvious that excludes the AI agents that are presented - although maybe additional criteria of realism will be found as research progresses..

Thus, what we see in most AI movies are plausible agents and largely achievable ultimately with enough scientific and engineering research.

The Robots are Insane

Irrespective of the success of the movie as an exciting, well-told story, or the plausibility of the AI agent, movies containing agents with 'minds' evoke interesting questions based on comparisons with human behavior. From this perspective, there is a notable fact about almost all of the AI agents in these movies: in varying degrees, the agents show abnormal behavior, from obsessive to pathologically insane. The main forms of deviant behavior are:

- Obsession with being loved: AI Artificial Intelligence, Making Mr Right, Electric Dreams

- Obsession with becoming human or at least physical: Demon Seed, Virtuosity, Wizard of Oz, Making Mr Right, DARYL, Iron Giant, Star Trek: The Motion Picture, Star Trek: Generations, Bicentennial Man

- Extreme behavior from irreconcilable conflicts (real life and complex society demands many different responses and actions humans. Sometimes these are hard or impossible to simultaneously satisfy, which may lead to robot psychopathology): HAL in 2001: A Space Odyssey.

- Megalomania: Virtuosity, Colossus - The Forbin Project

- Other obsessions: Dark Star

These are all common human behavior disorders, when taken to extremes. Other types of mental abnormality that afflict humans in varying degrees could or very occasionally do appear in the movies:

- Paranoid self-preservation: because humans could 'turn off the power', they are a threat, needing control, elimination, displacement, etc. This is a natural individual survival instinct enlarged beyond social control.

- Flawed or ungrounded reasoning: somewhat like Hamlet, an agent following a chain of reasoning that is divorced from real data, social conventions and feedback can come to odd conclusions. Alternatively, reasoning about material objects based on physical properties is likely to be stable, but what about reasoning about humans and behavior? Even as humans, we constantly misinterpret situations because we use incorrect assumptions or heuristics. AI agents will be no different, except some reasoning mechanisms may be beyond improvement, *e.g.* based solely on formal logic.

- Flawed perception or hallucinations: sensory data, particularly visual data, is complex, confounded with shape, position, illumination and sensor range and response. Humans resolve the confusion by using a combination of active perception, knowledge of the normal world and knowledge of specific objects. AI agents using unsound processes could make seriously incorrect conclusions about the external environment and thus behave oddly.

- Superiority complex: because of their greater computational speeds, broader perceptual ranges and mechanical strengths, AI agents might ultimately conclude that they are sufficiently superior that they need not treat humans (nor even other AI agents) properly.

Coming back to the cinema, these forms of abnormal behavior have had some but not a lot of exploration from the perspective of central characters in movies (*e.g.* Skynet in The Terminator is the cause of the main conflict but does not appear in the movie). So, we might see some more movies pursuing these aberrations.

Why then is insanity or extreme behavior so common?

People have always had anxieties arising from the uncertain future that a new technology might create. In this general sense, AI agents are no different than other technological devices, such as genetic engineering and nuclear power which also cause much justifiable anxiety. In those cases, the technology is already actively practiced, whereas AI technology is very far from being able to cause the problems that people worry about.

An AI agent makes our technological anxieties clearly visible, because the technology has the potential to displace not only our physical, but also our mental labor, and potentially even ourselves as the dominant agent on the planet. Given the success of mechanical devices, it is clear that people generally expect robotic devices to be stronger. Further, given the constantly increasing speed of computers, most general viewers would expect an AI agent to think faster than a human. Thus, having an AI agent in a movie allows us to explore our relationship with future AI agents, whether as master, equal or slave.

To give humans space to still be superior, movies credit humans with the ability to think better, clearer or more sensibly. Having an insane AI agent thus gives our poor human egos some boost, much like the 'mad scientist' caricature expresses general anxieties about the

power that comes from privileged knowledge and reassures the rest of us that "at least we're still sane". Alternatively, insanity gives AI agents a potential "Achilles heel" to exploit in conflicts. Hence we see AI agents with flawed reasoning of various sorts, such as not having real-world experience or common sense or not knowing something that only born, feeling, mortal humans would know. The technical reasons that lead to these shortcomings may be hard for the general public to understand. Insanity, on the other hand, is easily understandable and so can be more easily exploited in a story.

The scientific era has also subjected humans to an extended loss of status, largely leaving only our spiritual dignity intact (so far). People do debate whether animals possess consciousness and some spiritual side, but no mainstream religious or spiritual beliefs claim that mechanisms have souls. This implies that there should be a difference in behavior between humans and AI agents. Given the value placed on having a soul or spirit, this places AI agents in an inferior spiritual position. Although this does not imply insanity, it certainly implies "not quite human".

Robot Mental Healthcare

I am not deeply considering the question of whether real AI agents will go insane, especially since the sorts of AI agents depicted in these movies are many years away. But, it is interesting to speculate about what characteristics of AI might lead to insanity on the part of the agent.

In humans, insanity seems to arise because of deficient or defective perceptual, memory or reasoning mechanisms. As well as these physical disorders, mental impairment could arise from a variety of social causes such as isolation, inadequate or inappropriate socialization, sensory deprivation, torture, mental cruelty, antisocial companionship, lack of attention, contradictory obligations, hazardous circumstances, etc. AI agents are unlikely to be different. Also, because AI agents have different reasoning mechanisms, they are likely to manifest unexpected aberrant behavior as well, but the forms of this are harder to predict.

In the same way as human society has a variety of ways to cope with aberrant human behavior, we should anticipate movies exploring how aberrant AI agents might be controlled. Thus, as AI sophistication develops, we might see movies that investigate how human society might:

- Detect, prevent, control, or repair 'insane' AI agents (such as isolation, group therapy, 'psychotherapy', rehabilitation and destruction).
- Create an acceptance of and a way to act in irreconcilable conflict situations, perhaps by using a structured behavior priority system, such as Asimov's Three Laws of Robotics [3], to help resolve goal conflicts.
- Create an acceptance of agent 'death', or limiting the 'will' to survive.

- Engage AI agents in a 'society' that provides social feedback, behavior modification and constraints.
- Use 'drug' control (*e.g.* by producing modifiers to limit reasoning and perceptual abilities).

Future Films?

Is insanity what we can expect from any future movie having an intelligent, interactive AI-based agent? This is again an artistic issue, so it is hard to predict. Technological developments are likely to bring the potential of real AI agents closer. Moreover, the social issues that promote the use of AI agents as alternatives in humans in movies are unlikely to change. Although the popularity of science fiction fluctuates over time, the issues involved mean we will still see movies that explore the consequences of AI agents.

We commonly ascribe antisocial behaviors to "other people" in prejudices that we no longer (in theory) apply based on nationality, race, religion, gender, social class, etc. AI agents allow us to explore many of the same questions about aberrant human behavior explored by mainstream movies, in a context that allows us to ignore (or alternatively make explicit) questions of racism, sexism, etc. When an AI agent doesn't act properly, we could easily dismiss this with a sense of superiority; with humans we cannot or should not. By treating a nearly human AI badly, movies can comment on how we treat other people not quite like ourselves. It likely that we will see more AI agents in situations where humans typically appear, to explore our human concerns, anxieties, conflicts and histories.

On a more speculative note - why should we even assume or expect that real AI agents will behave in any manner intelligible to humans (Star Trek: The Motion Picture)? Even 'identical' twins diverge throughout life and human societies develop different and often hard to understand cultures. The AI agents' sensory, memory, reasoning and physiology systems will be quite different from ours, and all these shape their intellect. Thus, real AI agents could easily be incomprehensible. Given the many potential outcomes from the consequences of artificial sensing, motivation and reasoning, this seems like a fruitful theme for movie exploration as well as AI research.

As the majority of movies contain an unfavorable representation of AI agents, there might be problems for scientists investigating AI. If the horrors of nuclear war were well known in advance of the development of nuclear weapons, there would have been stronger social controls. I am not equating AI agents with nuclear weapons, but a sensitized general public might feel this way. Some AI scientists also strongly promote this nightmare. Raising these issues is correct and will help form a consensus about the allowable roles for AI agents in society. But, considering the 50-100 years likely before real agents appear that have the abilities of the agents in the movies, the movies and the concerns aroused may do much harm to AI research, even when that research is really just focussed on tools that enhance human abilities and activities rather than true AI agents. So, we might see some "backlash" movies.

There are a few recent movies with AI agents that are not obviously aberrant or insane. One example is the R2D2/C3PO team in the Star Wars series, but these AI agents do not aspire to greatness and equality, and we are disarmed by their comic cuteness or incompetence. A more interesting example is 2010: Odyssey Two, where the resurrected HAL agrees to a heroic self-sacrifice once it is fully informed of the situation and reasons. It is interesting that it is presented as making a sacrifice - why should it have a fear of 'death', when it can just be 'backed up' and restarted elsewhere? Perhaps this would be a consequence of needing some sort of survival drive to be an effective AI agent.

In spite of the many negative images of AI agents, the movies still inspire and excite people. It must be the excitement from a combination of the high-tech gee-whiz factor coupled with good adventure stories. As an AI researcher, I have also enjoyed the movies for both viewing pleasure and inspiration. Although I hope we see more positive representations of AI agents, I look forward to seeing the next AI agent, sane or otherwise.

Acknowledgements

I'd like to thank many people for their thoughts on this theme, but particularly Jim Bromer, Graeme Ritchie and Craig Robertson.

References

[1] AAAI's Science Fiction page, `http://www.aaai.org/Pathfinder/html/scifi.html`, March 28, 2002.

[2] About.com's Artificial Intelligence in Sci-Fi Movies, `http://scifimovies.about.com/library/weekly/aa013000a.htm` March 28, 2002.

[3] E. Seiler, J. H. Jenkins, Frequently Asked Questions about Isaac Asimov, `http://www.clark.net/pub/edseiler/WWW/asimov_FAQ.html` Oct 1, 2001.

[4] R. Fisher, "Representations of Artificial Intelligence in Cinema", `http://www.dai.ed.ac.uk/homes/rbf/AImovies.htm`, March 28, 2002.

[5] M. Hurt, "The University of Illinois's Cybercinema page", `http://www.english.uiuc.edu/cybercinema`, March 28, 2002.

[6] Per Schelde. Androids, Humanoids, and Other Science Fiction Monsters: Science and Soul in Science Fiction Films. New York Univ Press, June 1994

[7] Small Wonder Fan Club's list of TV and movies with robots and cyborgs. `http://smallwonder.hispeed.com/COC.html`, March 28, 2002.

[8] J.P. Telotte. Replications: A Robotic History of the Science Fiction Film. University of Illinois Press, 1995.

A Some Movies With Mindless AI Agents

A.P.E.X.	1994
Alphaville	1965
Assassin	1986 (TVM)
Aztec Mummy Vs. the Human Robot	1957
Battlestar Galactica	1978 (TVM)
The Black Hole	1979
Cherry 2000	1987
Chopping Mall/Killbots	1986
The Day The Earth Stood Still	1951
Dopey Dicks	1950
Eve of Destruction	1991
Forbidden Planet	1956
Futureworld	1976
Giant Robo	1991 (anime)
Hardware	1990
King Kong Escapes	1968 (animation)
Knights	1993
Lost in Space	1998
Mystery Science Theater 3000: The Movie	1996
The Outsider	1997
PatLabor, the Movie:	1989 (anime)
Red Planet	2000
Robot Carnival	1987 (animation)
Robot Jox	1991
Robot Wars	1993
Robot Monster	1953
Robotech: The Movie	1986 (animation)
Robotech II: The Sentinels	1986 (animation)
Runaway	1984
Saturn 3	1980
Silent Running	1971
Sleeper	1973
Space Truckers	1997
Sphere	1998
The Stepford Children	1987
The Stepford Wives	1974
Superman III	1983
Target Earth	1954

The Terminator	1984
Terminator 2: Judgement Day	1991
THX 1138	1970
TRON	1982
Wargames	1983
Westworld	1973

B Some Movies Having AI Agents With Minds

2001: A Space Odyssey	1968
2010: Odyssey Two	1984
AI - Artificial Intelligence	2001
And you thought your parents were weird	1991
*batteries not included	1987
Bicentennial Man	1999
Colossus: The Forbin Project	1969
Dark Star	1973
D.A.R.Y.L.	1985
Demon Seed	1977
Electric Dreams	1984
Evolver	1995
The Ghost in the Shell	1995 (anime)
Gog	1954
Heartbeeps	1981
The Invisible Boy	1957
The Iron Giant	1999
Knight Rider	1982 (TVM)
Knight Rider 2000	1991 (TVM)
Knight Rider 2010	1994 (TVM)
Making Mr. Right	1987
Matrix	1999
Max Headroom	1985 (TVM)
Metropolis	1926
Short Circuit	1986
Short Circuit 2	1988
Smart House	1999
Star Trek : First Contact	1996
Star Trek : Generations	1994
Star Trek : Insurrection	1998
Star Trek: The Motion Picture	1979
Star Wars	1977
Star Wars: The Empire Strikes Back	1980
Star Wars: Return of the Jedi	1983
Star Wars: The Phantom Menace	1999
Virtuosity	1995
The Wizard of Oz	1939