Research Statement

Manoj Ramanathan (manojramanathan.github.io)

Development of Customizable, Adaptable, Migratable AI platform for Intelligent Agents

Introduction

On a daily basis, humans interact and perform various tasks with several computing devices. Tasks once done manually or only on a desktop PC, laptop can be done on most mobile devices. These intelligent devices have taken different shapes and forms ranging from mobile phones to humanoid robots. With change in time and technological evolution, a paradigm shift can be observed in people's psychology and usage of various devices. Each of these devices are not readily accepted by the users. The embodiment of these devices and their interaction capabilities are defined by its users, device hardware limitations, application domains to provide a holistic and immersive interaction experience.

An integral component of such devices is the ability to understand the user's behavior using Artificial Intelligence modules. 'What is Artificial Intelligence (AI)?'. [1] provides a widely accepted definition that states AI as the ability of a machine to mimic "cognitive" functions of a human mind, such as "learning" and "problem solving". A very important aspect of any intelligent AI device or agent is its ability to perceive the operating environment and be able to cognitively understand and react to situations. These devices, referred to as "Intelligent Agents", can take any embodiment or shape. The intelligent agents can range from a simple conversational chatbot or web interface to more complex embodiments such as virtual entities (like virtual humans such as game characters, objects) or robots (humanoids and non-humanoids).

Ever since childhood, I have been fascinated by virtual characters (cartoons and game characters) and robots ("Vicky" in the TV series "Small Wonder"). Over time, with relevant research experience in AI and human computer or robot interaction (HCI/HRI), I have realised the possible capabilities and limitations that come with the development of these intelligent agents. They are expected to understand, react to situations, and proactively participate in an interaction if necessary. Based on the application domains, tasks to be completed, user preferences, the design of the complete AI pipeline and components are varied and difficult. My research focuses on developing AI modules that are:

- Migratable or Adaptable - Ability to adapt to different embodiments ranging from robots, virtual characters to simple chat interfaces.
- Customizable - Ability to be customized for different application domains.

Overall, I would like to begin with the development of AI modules for virtual characters (a good foundation) for specific industries such as gaming, management and financial sectors, healthcare, etc. The scope of this research can be further expanded to other embodiments and industries, thus truly making the AI more adaptable and customizable according to the user's needs. In the rest of the statement, I'll talk about my relevant research experience, motivations, and future scope of my research.

Research Experience

In my early years of research, during a discussion with my PhD supervisor on potential research topics, he introduced me to the field of computer vision, specifically, action recognition. The concept

of a device such as a computer being able to recognize actions from videos was really captivating. At this stage, the field of computer vision was entering a transition phase from hand-crafted features with machine learning algorithms to end-to-end trained deep learning neural network solutions. Due to this, I had the opportunity to work on both kinds of solutions, hand-crafted features and deep learning solutions.

During my PhD, I had worked on several possible modalities of action recognition data. Even though my main focus was on RGB videos, I had also worked on 3D (RGB-D videos), skeleton data for recognizing actions. To understand the field of action recognition and the various research challenges, it was essential to work on different types of data and get a broader perspective of the field. Due to this perspective, I published [2] focusing on research challenges in the field of action recognition and how various methods and available datasets adequately represent them. This helped me to understand the challenges in the field that paved the way to my doctoral thesis titled "Pose-Invariant action recognition for automated behavior analysis" [3]. Pose-invariant action recognition refers to the ability to recognize actions in a view-invariant, occlusion resistant manner. For this purpose, I had proposed and published [4], a mutually reinforcing pose and motion framework inspired by control system theory. The proposed features were invariant to human posture changes in videos. That is the action doesn't have to be done in a stand erect posture. This research was supported by National Research Foundation (NRF), Singapore, and A*STAR, Singapore.

A*STAR, Singapore is a leading research organization that works with several industrial partners and projects to develop innovative research solutions. Due to the industrial perspective, apart from just reliable recognition of actions, a lot of importance was provided on where and how the final action recognition model would be used. For instance, the action recognition model can form the backbone of an AI engine that tells if you are doing an exercise correctly. Given a reference exercise video, the AI engine would compare your exercise video and provide a detailed analysis of how correctly you are doing it. This was a big motivation for my future research. Apart from just developing end-to-end deep learning solutions for specific computer vision or natural language processing (NLP) problems, I was more curious in understanding how an overall AI engine operates, how the information from each deep learning model is processed and handled. A perfect way to explore this was provided in my post-doctoral research fellowship at Institute for Media Innovation, NTU, Singapore.

During the research fellowship, I had the opportunity to work on the AI engine platform of Nadine Social humanoid Robot [5, 6] and Virtual animation social assistant characters Megan and Nicole. In my initial year of research fellowship, my primary focus was on developing and integrating an action recognition model to the AI platform. During the course of this research, there were two main research understandings. Firstly, the variation in action classes that had to be recognized depends on application domains. As the AI platform was used for a social humanoid robot or virtual character, action classes were to be determined by social interactions and applications. Therefore, it was necessary to choose actions that have interaction significance such as if Nadine realises that you are making a phone call or putting on a jacket, she can use these actions as interaction cues to start a conversation. Secondly, to recognize such actions it was necessary to work on a multi-modal approach of action recognition that combines skeleton, 3D action data and contextual cues such as objects being used (from object recognition deep learning models). For this purpose, I had to work on each of these modalities separately and how to combine these in a framework. Apart from recognition, it was also necessary for me to integrate actions as a stimuli to the AI engine that controls Nadine. Thus, I was able to understand how an AI platform handles a single stimulus for its reactions. Please refer to [7] for details on the action recognition model developed for this purpose.

After the initial years of my research fellowship, I took over as the project lead of the complete AI platform that controls either social robot Nadine or virtual character Nicole/Megan. During this time, I was able to work on the complete AI engine and witness how it handles various stimuli, how it can be modified according to the agent (Robot or Virtual human), how it processes according to

various applications such as banking, healthcare, etc. Based on the observations, a customizable AI-empowered platform [5] was made and reworked for Nadine social robot. This AI platform is based on the "perception - processing - interaction" blocks. Being a social robot, Nadine has to understand social cues of the user and her environment and provide appropriate verbal and non-verbal responses. The main objectives of our design is to maintain human-like natural social behavior even in complex situations and be generic to handle any kind of data. Each layer consists of several sub-modules for specific tasks. Sub-modules can be added/ removed into each layer based on our requirement. Please refer [5] for further details. Apart from this, the research on developing a common AI platform for virtual humans and social robots is discussed in a SIGGRAPH ASIA course [8] co-authored by me. This became one of my main motivations for developing customizable and adaptable AI platforms that can work with any embodiments and application domains.

Since the AI platform consists of several stimuli, it meant that I worked with integration of several computer vision and NLP deep learning models. During this time, I was a research mentor to several graduate students and interns to develop and integrate several deep learning solutions to the AI platform such as face recognition, object recognition, chatbots, sentiment analysis, etc. This provided me a platform to interact, brainstorm, and develop solutions as a team and if necessary have a one-on-one conversation with researchers. In my view, teaching or mentoring is an art that involves the transfer of knowledge between both parties. In an ever-changing world, as much as a student, the teacher must also be willing to accept and learn new knowledge from all sources. These interactions with students allowed me to hone my teaching skills as well.

Nadine social robot is a worldwide famous humanoid robot, which has made several appearances in various events and is involved in several Human-Robot Interaction (HRI) experiments with different applications such as in banking or insurance (with AIA Singapore Pte. Ltd), healthcare or elderly care (with nursing homes for dementia). For each of these applications, the customizable AI platform had to be changed to consider the various tasks Nadine would be performing, her operating environment etc. For instance, the chatbot data has to be modified, the affective engine controlling emotions of Nadine etc have to be modified and customized to generate socially acceptable AI-empowered behaviors. I was involved as the lead researcher and was involved in writing the research proposal for these projects. Effect of Nadine's work at AIA Singapore was studied as an HRI experiment in [9] by us and was jointly funded by AIA Singapore. Currently, Nadine is working as elderly care social companion in a dementia care nursing home. Apart from these, Nadine has also been interacting with customers at several company conferences, art science museum etc. During these conferences, it is necessary to modify the AI platform so that she can answer queries according various customer from different fields including, arts, healthcare, financial sector etc. I have worked on successfully modifying and customizing Nadine's AI module for these events. All work on Nadine social robot was supported by Being Together Center, a research collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill.

After my research fellowship, I have been working as an AI research scientist in a startup Dex-Lab Pt. Ltd, Singapore (web: https://www.dex-lab.com) that develops virtual characters and Lifestyle humanoid robots (intelligent agents) for commercial applications. As a way to explore the domain of industry-related research and to know what customers or clients require from such intelligent agents, I made a switch to the startup. The role was similar to my post-doctoral research fellowship to develop customizable AI platform for intelligent agents for commercial applications such as office receptionist, mall concierge, healthcare assistant, and teacher or educator. Similar to my previous work, the virtual character (Chloe) and humanoid robot (Dexie) have a common AI platform making it migratable . Based on my previous experience, I completely reworked a new customizable platform that can be adapted to both embodiments and different application domains mentioned above.

Both AI platforms were customizable to work with virtual characters (Megan, Nicole, Chloe). For this, it was essential for me to work on Graphics and Unity 3D (game engine) programming. This was

a completely new field for me and was very interesting and challenging. The last essential component of the AI platform was to generate socially accepted behaviors during the interactions based on the various stimuli being processed. These behaviors had to be modified according to the hardware capabilities (Robot / Virtual character - embodiment change), application domain, etc. For these virtual characters, these behaviors had to be rendered and programmed into Unity 3D. One of my primary responsibilities was to develop and program in Unity 3D to control the virtual characters. This allowed me to understand the complexities of rendering, graphics and be able to add functionalities accordingly. During my research fellowship, I was editorial assistant of The Visual Computer (Impact factor: 1.456), a leading journal in the field of computer graphics, that publishes on research fields of capturing, recognizing, modelling, analysing and generating shapes and images. It includes image understanding, machine learning for graphics and 3D fabrication. The stint as an editorial assistant was useful in two ways. Firstly, I was able to understand the concepts of an editorial journal and different stages of publication. Also, it involved preparing the journal publication issues and writing the preface for them. Secondly, I was able to read the current state of the art on research topics such as machine learning, deep learning, AI and graphics that several researchers worldwide are working on. This helped me strengthen my understanding of graphics and AI related research as well. My fascination for cartoons, the experience of working on graphics and virtual humans have been a driving force for me. This is one of the main motivations behind my research interests to develop virtual characters (human and non-human embodiment) for several industries such as gaming, banking, insurance, management sectors, healthcare, etc that can be AI empowered.

Future Research Directions

Developing a customizable, adaptable, and migratable AI platform to work with any intelligent agent has been the main research focus that has driven me so far. During such a development, my work would explore the following research themes and topics:

Understanding and generation of speechless or non-verbal interaction cues - Most AI modules can be broadly considered to be under three categories or pillars, namely, Audio (Speech or Music), Computer Vision, and Natural Language Processing. But an important component of interaction is the ability to understand and portray non-verbal behaviors. As a preliminary research, I studied various possible speechless interaction cues such as eye gaze, empathy or affective engine, reading, social media presence, etc in social robots. This research has been agreed to be published as a book chapter [10] (I was also one of the editors of the book). An interesting research that I would like to pursue is how the AI platform could be customized to automatically understand and generate non-verbal interaction cues for any of the intelligent agents.

Generation of AI empowered natural behaviors - As mentioned earlier, any AI platform can be considered to have 3 blocks, namely, perception, cognition, and action. In these blocks, most deep learning models operate to perceive the state of the user and environmental stimuli. Based on the application domains, hardware capabilities (embodiment) and tasks to be achieved using the intelligent agent, the set of possible behaviors are already predefined. In cognition and action blocks, based on the perceived states and predefined set of behaviors, a socially acceptable and appropriate behavior is chosen, which is portrayed by the agent. One of the research avenues would be to create an end-to-end training of deep learning model that can generate natural behaviors based on the intelligent agent capabilities, application domains and perceived states. This would be a long term research project as it is a challenging topic with several scenarios and use cases. Also, data collection and HCI experiments have to be conducted for each of these cases.

Financial and Management domain application for Intelligent agents - In my career, so far, I had the opportunity to work with social humanoid robots and virtual humans in domains such as banking,

insurance, office assistant and receptionist etc. All of these applications allowed me to understand several nuances in financial sectors. To create a more intelligent agent, it would be necessary for the agent to be equipped with predictive analysis using the various financial sector data. Inclusion of such a model would enhance the agent's ability to answer and provide meaningful solutions to customers in these sectors. In my research, I would like to explore this and see how it can be integrated to the agent's socially accepted behavior. Another interesting application would training of new bankers and management students using a virtual human or agent. Development of virtual humans that can various roles like customer, banker, other officials etc and various personality traits is yet to be explored. It would be interesting to conduct user studies and possibly have fully developed product out of this study.

Experiments with Different embodiments and Application Domains - Up until now, most of the intelligent agents that I have worked on have a humanoid embodiment (Robot or virtual human). But the objective of a migratable AI platform is that it is capable of handling other types of embodiments such as non-human virtual characters or robots, web-based conversational chatbots, voice assistants, etc. An interesting research would be to see how AI platform will be able to migrate to these embodiments and still be able to generate proper verbal and non-verbal interaction cues. The work in Dex-Lab Pt Ltd, Singapore allowed me to look into several social applications for virtual humans and humanoid robots such as healthcare assistant, mall concierge, office receptionist, etc. Due to deadline restrictions and focus on developing task requirements of clients, it was not possible to conduct HCI user acceptance experiments for these applications. Conducting these experiments could validate and bring out several research insights required for developing a truly customizable and adaptable AI platform, which is why I would like to delve into this research avenue further.

Experiments with Non-social applications (like Gaming) - Most of my research experience so far has been on agents that have been used in social interactions and applications. In some ways, this is a limitation to explore several components of the AI platform. For instance, consider the development of an empathetic AI (to perceive and portray human emotions). In social applications such as healthcare assistant or office receptionist, the agent cannot exhibit negative emotions or behavior. But in a non-social application like gaming, we could let the AI-empowered virtual characters explore the complete range of empathetic AI spectrum. In the gaming industry, it is easy to create virtual characters (both human and non-human) that can be empowered by the proposed AI platform. The virtual character created can also be considered as a precursor or simulated version of the final user interface to be developed. In terms of gameplay, a separate research could be on using AI to generate storyboards dynamically during the game. Exploration of such non-social applications would open research avenues as it requires generation and collection of data to train the models and conduct HCI experiments to validate the hypothesis. It would also open the possibility of creating immersive games using VR/AR technologies. Such research would interest gaming company giants who aim to make more immersive and realistic gameplays and might fund the projects.

Miscellaneous research - The above-mentioned research themes could also be expanded into some research applications such as the ability to recreate virtual avatars (human and non-human) that mimic natural verbal and non-verbal behavior to create an immersive storytelling experience by using visual recreation of story, empathy, and intonations during narrations, etc.

References

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. USA: Prentice Hall Press, 2009.

[2] M. Ramanathan, W.-Y. Yau, and E. K. Teoh, "Human action recognition with video data: Research and evaluation challenges," *IEEE Trans. on Human Machine Systems*, vol. 44, no. 5, pp. 650 – 663, October 2014.

[3] M. Ramanathan, "Pose-invariant action recognition for automated behaviour analysis," Ph.D. dissertation, Nanyang Technological University, Singapore, 2017.

[4] M. Ramanathan, W.-Y. Yau, E. K. Teoh, and N. M. Thalmann, "Mutually Reinforcing Motion-Pose Framework for Pose-Invariant Action Recognition," *Intl. Journal of Biometrics*, vol. 11, no. 2, pp. 113 – 147, March 2019.

[5] M. Ramanathan, N. Mishra, and N. M. Thalmann, "Nadine Humanoid Social Robotics Platform," in *Computer Graphics International (CGI)*, ser. Advances in Computer Graphics, Part of LNCS book series, M. Gavrilova, J. Zhang, N. M. Thalmann, E. Hitzer, and H. Ishikawa, Eds., vol. 11542. Springer, Cham, June 2019, pp. 490 – 496.

[6] IMI NTU. (2020) IMI NTU Singapore, YouTube page. [Online]. Available: https://www.youtube.com/user/IMINTUsg/featured

[7] M. Ramanathan, J. Kochanowicz, and N. M. Thalmann, "Combining Pose-Invariant Kinematic Features and Object Context Features for RGB-D Action Recognition," *Intl. Journal of Machine Learning and Computing*, vol. 9, no. 1, pp. 44 – 50, February 2019.

[8] D. Thalmann, N. M. Thalmann, and M. Ramanathan, "Real Humans with Virtual Humans and Social Robots Interactions (HCI)," in *SIGGRAPH Asia 2017 Courses*, ser. SA '17. New York, NY, USA: ACM, 2017, pp. 15:1–15:221. [Online]. Available: http://doi.acm.org/10.1145/3134472.3134513

[9] N. Mishra, M. Ramanathan, R. Satapathy, E. Cambria, and N. M. Thalmann, "Can a Humanoid Robot be part of the Organizational Work Force? A User Study leveraging on Sentiment Analysis," in *28th IEEE Intl. Conf. on Robot and Human Interactive Communication (RO-MAN)*. IEEE, October 2019.

[10] M. Ramanathan, R. Satapathy, and N. M. Thalmann, "Survey of Speechless Interaction Techniques in Social Robotics," in *Intelligent Scene Modelling and Human-Computer Interaction, Human Computer Interaction Series*, N. M. Thalmann, J. J. Zhang, M. Ramanathan, and D. Thalmann, Eds. Springer, Cham, 2021, pp. 241–257.