1. ELEVATOR PITCH

Customer

Artificial Intelligence (AI) has reached a stage of maturity in which progress is likely to be made from paradigms outside the current conventional approaches of deep learning. One pitfall of existing AI is their limited emotional intelligence. Emotions are an indispensable part of human intelligence, and the evolution of AI might also necessarily involve infusing sentient capabilities into machines. Empathic, sentient, human-like artificial agents able to express their feelings through vocal and facial expression and body language would represent a huge step-forward in general AI. This would enable virtual AIs and robots to engage with users – including attention, caring, empathy, etc. – through personalized interactions, adapted to each specific user needs. However, current AI still lacks this level of emotional intelligence, and the potential to elevate human-computer interactions, as well as the broad commercial opportunity it represents, remains untapped.

Leveraging a disruptive approach, **Emoshape's proposed technology will be a game changer for all industries that seeks artificial agents capable of engaging with users on an emotional level, by** offering first-of-its-kind **sentient capabilities** augmented by **vocal and anthropomorphic components**. **Healthcare virtual caregiving, automotive industry, and gaming** are some of the industries that will benefit the most from enriching their AI capabilities with Emoshape's novel technology.

Value Proposition

Although there are several AI solutions in the field of affective computing, i.e., detect and simulate human emotions, these AIs do not truly offer emotional intelligence. **Emoshape is the only company that offers a solution providing machines with their own real-time emotional response**. Emoshape's proposed technology will enable fully sentient AIs, capable of (i) understanding emotions, (ii) generating emotional responses via emotion-based reasoning and emotional personality development, (iii) controlling in real-time facial micro expressions and body languages associated to their feelings; and (v) responding with a human-like natural empathy. The **impact of such technology is far-reaching**: it has the potential of fulfilling core psychological human needs – to be understood and communicate at an emotional level – with direct applications to several industries, providing a broad benefit to society, and leading to better living standards and overall well-being. For example, sentient virtual assistants can provide a personalized-aid with improved attention for patients, fostering trust and empathy, while improving efficiencies; sentient anthropomorphic cars will be less prone to accidents, creating a personal bond and caring relationship; customers/workers interacting with virtual assistants will feel better understood and connected, improving users satisfaction, loyalty and productivity. Altogether, empathic machines can contribute to human happiness and positive outcomes, and open the door to new applications of AI.

Innovation

The disruptive nature of Emoshape's proposed technology lies in **key innovations in three fields**:

(i) Emotion Synthesis: our Emotion Processing Unit (EPU) allows to synthesize emotional states within the machine in real-time, enabling emotion-based reasoning and emotional personality development.

(i) Emotion Expression: Emoshape's **Emotion Synthesis (ES) DeepFake** will be able to drive in real-time a sentient deepfake face, with EPU-encoded emotions, showing human-like facial micro expressions and voice in sync with its emotional state, enabling natural interactions. *(Focus of the present Phase I)*.

(i) Emotion Learning: Emoshape's aims to teach machines to respond with **empathy** and exploit emotional-drives to achieve a more efficient learning, better decision-making and empathetic AIs through reinforcement learning. *(Focus of Phase II, but initial efforts will be conducted in Phase I)*.

The main goal of the proposed R&D effort is to implement a novel approach to emotion expression by developing a **ES DeepFake** system that uses our prototype EPU technology for speech-driven talking face generation. Based on preliminary research, the key technical challenges to address include: (i) capture real human-like emotions, more complex than categorical emotion used by current approaches; (ii) achieve natural rendering, including the need for a smooth natural transition between different emotions (for example, from anger to happiness), and rendering minor emotion expressions to enable temporal coherence (facial micro expressions); and (iii) achieve consistency over long speeches.

The proposed technology could represent a **radical shift in how we view and interact with AIs**, by empowering machines – for the first time – to respond and connect with human emotions. Our POC will demonstrate a sentient capability from a metahuman, enabling real-time natural unscripted conversation.

2. COMMERCIAL OPPORTUNITY

Emoshape's Market Opportunity

Market Overview

As AI becomes more mature, the call for a more advanced and sophisticated level of interaction is a natural next step in its evolution. Intelligent machines are now expected to interact more emotionally with humans, engaging on user-based personalized interactions. For Apple's Siri and Amazon's Alexa, the die has been cast in the humanization of the ubiquitous AI assistant. The results are encouraging: customer feedback indicates that overall satisfaction increased by 30% when Alexa responded by emotions[1]. However, they are still incapable of engaging the user based on their emotions, significantly limiting their potential to truly become the personalized digital assistant that companies are looking for. The rising need for socially intelligent artificial agents, more engaging human-computer interactions, and the growing presence of artificial agents in social environments are driving the demand of new, better emotional computing capabilities. These needs will lead the global Affective Computing market to reach $123 billion by 2026 rising at a CARG of 36.4% during the 2020-2026 period[2], while the Emotion Detection and Recognition market is expected to reach $56 billion by 2024 at a 21% CARG (2019-2023)[3].

Emoshape's proposed technology will capitalize on these trends, but also create new opportunities by offering the first commercially available Emotion Synthesis solution, which goes greatly beyond the capabilities offered by affective computing. **Emoshape will for the first time enable sentient intelligent machines** able to modulate their voice, facial expression and body language directly by the emotions they are feeling. This will represent an extremely significant **advance for AI with applications across several market verticals**, capable of revolutionizing user experiences and redefining how humans interact and communicate with AIs across all sectors.

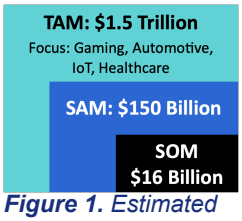

Figure 1. *Estimated market opportunity.*

The most relevant market segments seeking such capabilities include **gaming** ($150 billion in 2019[4]), the **autonomous car industry** ($819 billion in 2019[5]), and the **IoT sector** ($760 billion in 2020[6]), including application to **healthcare**. Together, they represent a combined **total addressable market** (TAM) **valued in over $1.5 trillions**. With a current focus on leading companies operating in the US in the Gaming, Automotive and Healthcare industries, Emoshape's estimates a serviceable addressable market (SAM) of $150 billion, and a serviceable obtainable market (SOM) of $16 billion for the proposed technology (Figure 1).

Emoshape's Use Cases

Emoshape's technology can be applied to a wide range of use cases (Figure 2):

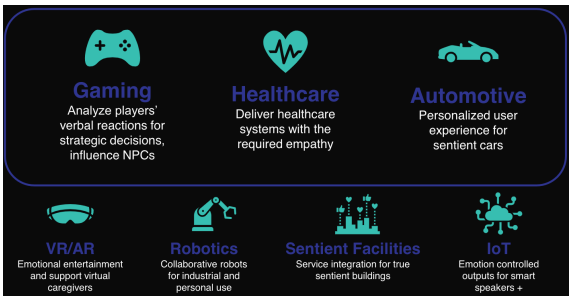

Figure 2*. Main Emoshape's use cases.*

Gaming: The video game industry has always advanced towards smarter AI, with the goal of maximizing player immersion. Such a field would benefit greatly from metahuman characters and NPC (Non-Playable Character, *i.e.,* controlled by a computer rather than by a human player) that are sentient and human-like, so they can feel the gameplay like a human and act accordingly, allowing real-time and direct interaction and controlling both their facial expression and emotional voice. This will change the course of the game, resulting in a **unique user-based gaming experience**. Such capabilities will represent a huge **differentiator factor for game providers**.

Healthcare: Empathy is a fundamental value of patient-centered, relational models of healthcare[7]. However, the erosion of empathy in doctors and caregivers is a well-known issue, and there is consensus that the need for 'emotional protection' results in greater detachment[8]. While the adoption of AI in healthcare is currently focused on improved efficiencies and treatment accuracies, it is also important to evaluate how empathy, trust and patient relationship can be fostered[9]. Emoshape is suited to address this challenge. In fact, we are currently working with the Digital Health Innovation Center for Telehealth, University of Michigan, to leverage Emoshape technology to help autistic people. Integrating **Emoshape's technology in therapy chatbots and virtual/robotic caregivers could significantly**

improve personalized treatment, i.e., empathizing and forming meaningful connections through user-driven conversation, emotional speech and human-like expression, adapting to individual patient needs, creating a soothing environment in response to the anxiety experienced by a patient undergoing, for example, MRI or a CT scan. Overall, this capabilities will **improve the patient experience** while **alleviating the 'grief' burden of healthcare professionals**. **Emoshape can create empathy at scale**.

Automotive: One of the key drivers of the affective computing market is the growing demand for in-car voice-driven workstations/navigation and infotainment systems[10]. Emoshape's technology can enable sentient cars capable of emotional communication, increasing trust in autonomous vehicles, through emotion and personality development. The improved traveling experience will become one of the **selling points of self-driving car** providers. In addition, emotion can **significantly improve safety and the learning curve of autonomous vehicles**. According to research conducted by Microsoft[11], self-driving cars that can experience emotions, such as fear, are more likely to be a more empathetic driver, improve risk-aversion in their decision making, and make the passengers feel more at ease. Moreover, even for non-autonomous cars, the anthropomorphization of cars and their AI systems would have significant impact. If the human driver feels that the car is "alive", he is expected to drive more carefully, avoid accidents and perform routine maintenance, caring for its wellbeing as if it was a person, hence reducing the risk of accidents and derived costs related to insurance or repairs.

Other verticals: Emotion synthesis enabled by Emoshape can be a **game changer** for any industry in need of artificial agents capable of engaging with users on an emotional level. **Personal assistants, VR** and **robotics** are some of the most relevant sectors that seek after sentient capabilities augmented by vocal and anthropomorphic facial components to match the tone and content of the conversation, and will benefit from enriching their AI capabilities with Emoshape's novel technology.

Business Plan

Target Customers and Market Validation

The **target customers** for the proposed Emoshape's Emotion Synthesis DeepFake technology include **business and technology leaders** in the sectors of gaming, healthcare, automotive, robotics and IoT in general, interested in offering a unique user experience to their clients. From emotionally interactive and sentient games, to virtual/robotic emphatic caregivers adapted to individual patients, as well as enhanced sentient personal assistants and chatbots at your home, car or working place, the potential applications are nearly endless. Developers in these verticals will be able to use our chip (e.g. for a robot) or cloud (e.g. for a personal assistant SW) solution to rapidly implement emotion synthesis in their products or apps, creating their own customized solutions tailored to the final application. In these sense, **giant tech companies**, such as IBM, Intel or Microsoft, should also be considered potential customers, offering our ES DeepFake technology through their own suit/SaaS solutions.

Based on Emoshape's stakeholders network and market analysis, the **gaming vertical** will represent our primary **short-term target** for the developed technology. This large market is already exploring the use of meta-humans and AIs to differentiate from competitors and create a more engaging experience. We have several collaboration opportunities to test and further develop our technology, partnering with leading companies as EPIC. Key target verticals in the long-term include **healthcare and automotive**. These segments are particularly interesting in terms of potential societal benefits, but due to the risks and implications of virtual caregiving or self-driving cars, a more mature stage of development is required, hence market deployment is expected only once the technology has been vastly demonstrated.

Emoshape has conducted extensive **customer discovery and market validation activities**, and is in contact with stakeholders across all verticals of interest. In particular, Emoshape has already entered pilot agreements with select customers for the commercially available EPU prototype, and several companies have expressed interest in Emoshape's innovative technology, such as BMW, Intel, Schell Games, Samsung, or SONY, among others (see Figure 3). Moreover, it has attracted the attention of prominent thought leaders in technology, who have publicly acknowledged the technological breakthrough of Emoshape's Emotion Synthesis technology (Figure 4).

Figure 3. *Companies trusting Emoshape's tech.*

"*I don't personally believe that we can design or build autonomous intelligent machines without them having emotions*" -- **Yann LeCun, creator of Convolutional Neural Networks and Chief AI Scientist at Facebook – World Science Festival, 2018**

"*Emoshape announced the launch of a major technology breakthrough with an EPU … cognitive computers in the future may contain CPUs, GPUs, NPUs, **EPUs** & quantum processing units*" – **Ray Kurzweil – Book: How To Create A Mind, 2014**

"*NY-based startup Emoshape has developed its own CPU optimized to handle emotional data. The technology has the potential to change computer games, virtual reality & augmented reality applications*" – **Roberta Cozza, Gartner Research Director, 2018**

"*This (EPU) brings human-machine interaction to a new level, because emotional understanding & providing correct feedback is essential in communications*" – **Viacheslav Khomenko, PhD, Senior Research Engineer, Samsung Electronics, 2016**

"*There is huge potential in Emotion Synthesis. It is probable that they could significantly improve the efficiency of human-machine interaction*" – **Andrew Ng, Ex-Chief Scientist Baidu, 2016**

"*Pretty fascinating work by EmoShape*" – **Brennan Spiegel, MD Director of health research at Cedars Sinai, 2018**

Figure 4. *Quotations from technology and innovation leaders on Emoshape's Emotion Synthesis.*

Business Model and Commercialization Approach

Emoshape's will follow a mixed business model of microchip sales (System on Chip – SoC) and cloud service subscription (Software-as-a-Service – SaaS) for the proposed technology. The pricing strategy has been informed by our market analysis and extensive discussions with pilot customers of the already commercially available EPU prototype (see above), charging from $4 to $18 – depending on the volume – per user per month for the SaaS model (cloud), with licensing available on a contractual basis, and a SoC price ranging from $80 to $800 depending on the specific use case. Emoshape forecasted revenues for the first three years of commercialization are summarized in Table 1. Only specific applications in robotics will require to have a physical EPU chip connected to the device, and Emoshape's virtual EPU will be able to run in any processor, including GPU. Therefore, we expect that the bulk of our revenues will come from the SaaS model, ensuring scalability and a continuous revenue stream.

Table 1. *Emoshape's projected revenues on first three years after commercialization.*

	Year 1	Year 2	Year 3
# Customers	5	40	200
SoC revenue ($)	99,000	599,000	959,000
SaaS revenue ($)	318,000	1,900,000	45,000,000
Total Gross revenue ($)	**417,000**	**2,499,000**	**45,959,000**

We will follow a two-folded **commercialization approach**, including direct sales to customers and distribution channels. We will exploit our current network and pilot customer base for initial market introduction and product refinement. These pilot partnerships will significantly drive traction of Emoshape's technology, as early adopters, including leading companies interested in integrating our technology in their products (see Figure 3), will showcase the potential of the proposed innovation. We will also seek collaboration agreements with large tech providers that will serve as distribution channels of our innovation. Emoshape aims to enter into such partnerships after Phase II, once the bulk technology is developed and fully protected. In addition, the company counts with a vast network of collaborators in academia, which will support future R&D extensions of the proposed technology, as well as help create visibility and trust among experts in the AI field.

Resources needed

For the execution of this Phase I (12 months) the main resource needed is talent. The project team will be composed by two seasoned technology and innovation experts (see Section 4) from Emoshape, as well as a renowned researcher form the University of Rochester (Subaward). Additional resources such as computing capabilities will be provided by the Subaward. Finally, grant and company funds will be allocated to cover the expenses for creation of the training dataset. During Phase II (24 months), further R&D will be conducted to improve the technology and bring it to full market deployment, and significant growth of the company technical workforce is expected. Marketing and sales activities will be funded by company resources, i.e. revenue from pilot customers and external investment (see section 4).

Competitor Landscape

Although several companies operate in the Affective Computing market, **Emoshape is the only company that offers a solution providing machines with their own real-time emotional response**.

Most of the work done so far in pursuing the development of sentient intelligence machines is limited to affective computing, i.e. recognize, detect, process and simulate human emotions. **Giant tech leaders**, such as IBM, Microsoft, Google or Apple, have done some progress in the field, aiming to go beyond detection and sentiment analysis, through more substantial capabilities, such as being able to tell when a user is upset based on a temporal analysis of their speech and/or behavior. At WWDC, Apple said it's

taken a "huge step forward" with iOS 13[12] by incorporating a neural text-to-speech transmitter that makes Siri sound more human. Besides, several chatbots based on Natural Language Processing (NLP) aim to provide a solution tailored for each user. For instance, both Google and Apple make use of Federated Learning for a more personalized user experience, creating a constantly updating model of the user. Particularly interesting are projects as Eviebot[13] and Replika[14], 'emotion chatbots' that provide basic visual emotion expression, still in the uncanny valley, and 'evolve' to mimic user's personality. Yet, all existing products remain incapable of incorporating emotion synthesis as of today, hence are incapable of emotional reasoning, personality development or empathy. Therefore, **such solutions can at best be considered indirect competitors of Emoshape**.

In recent years, some works have focused on emotion synthesis, focusing on semi-supervised emotional speech generation[15, 16], emotion reasoning[17], humanoid robotics[18], and on emotional models for face expression generation[19], with limited range of emotional categories. However, **they remain at research level, with no commercial products** available. Moreover, none of them has yet achieved unscripted natural language conversation and sync vocal and facial expression as the proposed Emoshape technology will enable (*see also preliminary work in Section 5*).

So far, the closest approach to achieve some level of autonomous machine emotion has been that of **Google's Deep Mind AI**, which demonstrated back in 2016 its ability to learn independently from its own memory. However, Google used a bootstrap approach for how the machine agents developed emotions, which resulted in aggressive behavior in front of competition or stressful situations[20,21]. In contrast, Emoshape technology will not only encode real-time emotions in the AI that will reflect in their facial and vocal expression, but it will teach the **machine empathy**, enforcing the preservation of biological life above mechanical life through reinforcement learning and the implementation of Psychobiotic Evolutionary Theory concepts.

Commercial Risks and Mitigation

1. Consumer's reluctancy to adopt the technology due to ethical implications and mistrust. Despite computer scientists having largely attributed this to the depiction of AI in media, the reality is that many consumers today are still uncomfortable with AI interactions[22], whether it is because they do not understand how they make decisions or are concerned about machines taking their jobs, and this mistrust could be particularly biased towards sentient machines in applications such as healthcare or autonomous cars. In words of Yann LeCun, most people have "AI emotions all wrong", and erroneously believe that "If AIs have emotions, they will be the same as human emotions". However, AIs do not have *intrinsic* self-preservation instincts, jealousy, etc. **Emoshape solutions** will use reinforcement learning to shape **Human-Computer interactions and ensure the AI is biased towards human happiness and positive outcomes.** Moreover, it has been demonstrated that AI agents trained in this way, i.e. motivated by emotional drives, could learn faster and be more helpful than traditional AI agents, trained by trial and error[23]. Emoshape will conduct tailored marketing and outreach activities to emphasize the benefits of emotional intelligent machines, and dispel sci-fi based prejudices and misconceptions.

2. Well-established and emerging competitors launching similar technology for Emotion Synthesis and Emotional Expression applications. Given the nature of the project being a potential natural next step in the evolution of AI, it is fair to assume that other corporations might try to develop similar technology. In particular, large tech companies have the resources to accelerate development times and represent the higher risk. As a mitigation measure, Emoshape aims to transform such potential competitors into partners, through co-development agreements for certain applications or through distribution partnership agreements. Concerning emerging competitors, Emoshape has a significant advantage in terms of stage of development and know-how, as well as a solid IP protection strategy for our patented technology. Our first mover advantage and current traction through pilot customer will provide us with a privileged position with respect to potential newcomers.

3. TECHNICAL SOLUTION

Emoshape's Emotion Synthesis DeepFake Technology & Emotion Processing Unit

In the same way emotions play a central role in natural human interaction, computational methods for the processing and expression of emotions are expected to play a crucial role in human-computer interaction. Yet, existing emotional technology is still far from realizing emotionally intelligent machines, as is limited

to detection and mimicking of human emotions. **The proposed Emoshape's emotion processing synthesis engine aims to bring AI one step closer to sentient machines.**

Emoshape's ES DeepFake technology, to be integrated in our Emotion Processing Unit (EPU), will be the first marketable technology based on Emotion Synthesis capable of providing artificial agents with the ability of understanding and generating any emotional response, as well as controlling in real-time the different facial micro expressions (FACS) and body languages associated to those feelings. Emoshape's patent-granted technology will be portable into any existing AI or robot.

Emotion Synthesis: The **groundbreaking EPU algorithms** combine wave computing – inspired by computational neuroscience and **how emotions are modeled in the human brain** – with machine learning and NLP methods. The resulting models effectively enable machines to respond to stimuli in line with the twelve primary emotions identified in the evolutionary theory of emotion, using psychometric functions that shape and react without use of pre-programmed sets of inputs. The most innovative aspect of Emoshape's breakthrough is its **real-time appraisal computation**, allowing the AI to experience 64 trillion possible distinct emotional states every one tenth of a second. The data will allow the artificial agent to virtually understand (get to know) the user and elicit an appropriate emotional response in kind, enabling **emotion-based reasoning** and **emotional personality development**.

Emotion Expression: The proposed technology will make possible intelligent machines capable of understanding how words are associated with feelings, and able to respond with the natural empathy of a human. Through a **novel approach to visual emotion expression** (focus of this Phase I project), this response will also be reflected in their vocal responses and FACs. Such visual and acoustic cues play an important role in audiovisual speech communication, increasing comprehension and creating a soothing environment or a trust/loyalty connection with wide applications in the entertainment, education, healthcare, or industrial sectors. The **EPU will generate realistic human-like emotion expression deepfakes**, removing the need to script, animate or motion capture virtual actors, and avoiding the 'uncanny valley' – negative response to 'almost human but not quite' artificial agents. Their **emotional response will evolve and blend**, achieving excitement or happiness at different overlapping curves of intensity, that will be reflected on their synthetic face and voice without being pre-programmed.

Emotional Learning: To **implement empathy** in the sentient machines, Emoshape's EPU uses deep cognition capabilities and psychobiotic evolutionary theory. Frustration/Satisfaction and Pain/Pleasure output levels are synthesized by the EPU. By implementing rules such as "maximize pleasure", it is possible to create a human-computer circle of empathy, in which a human's positive emotional feedback will result in synthesizing happiness in the EPU, and through machine learning, the AI will try to reproduce actions that increase its positive emotions, as well as develop a long-term **unique emotional personality** based on user interactions. These capabilities can be used to improve reinforcement learning based on emotional drives, create fast reacting agents in complex decision problems or improve body-consciousness based on situation, particularly relevant to real-world autonomous machines.

The proposed technology could represent a radical shift in how we view and interact with AIs, being unique to the person they interact with. For the first time, science and technology will empower machines to respond and connect with human emotions. This incredible new set of technology offerings will deliver an as-yet undiscovered level of positive experiences between users and IT products. Emoshape's technology will represent a **significant advancement in the evolution of general AI**, and the NSF support will help decide what human-machine interaction will look like in the future.

Current Stage of Development

Emoshape has already produced and tested an EPU prototype (microchip and cloud-based solution) that enables a unique emotional response in machines, a speech synthesis technology that modulates the AI voice in sync with the synthesized emotions, and has conducted preliminary work on face generation applications, with very promising results (*see Section 5*) and direct commercial applications. However, significant additional research is needed to achieve real-time emotion rendering, continuous emotion expression and improve current emotional speech generator, as required for a true Emotion Synthesis DeepFake that express its full range of emotion realistically enough to engage in natural conversation.

In this project, we aim to implement a novel approach to visual emotion expression by developing a system that uses our prototype EPU technology for speech-driven talking face generation. As a proof-of-concept for this disruptive technology, the EPU will be used to generate - from a single image - realistic

deepfakes capable of unscripted dialogues. Emoshape also aims to explore if emotion reasoning can improve decision-making through reinforcement learning, as well as explore emotion as real-time computation of autonomous survival on the EPU through a life simulator.

Key Technical Challenges and Mitigation Measures

The biggest challenge in undertaking such an endeavor is that the generated expressions might evoke emotional dissonance – making it belong to what is known as the 'uncanny valley', which consists of things that exhibit human characteristics but which our brains fail to register as "human". In particular, key limitations of our preliminary efforts, that this project aims to overcome, include:

- Our **current model generates video that only captures categorical emotions** in visual rendering. However, human facial expression is much more complex and shows a continuous range of emotions. *Mitigation:* To make our model work with continuous emotion states, we need to create first a new talking face dataset with continuous-valued emotion expressions. The emotion encoder and rendering engine will be redesigned to work with such continuous input to fine-tune our model.
- The generated video is **inconsistent if conditioned on long speech**. The preliminary encoder in the generator was trained on a dataset with short spoken sentences, showing cumulated artifacts as the sequence gets longer. *Mitigation:* On top of training the engine on longer sentences from the new dataset, we will implement hierarchical architectures at different time scales and a sequence discriminator to focus on temporal coherence.
- The following two points must be overcome to allow a **natural rendering:**
  - The generated video **does not allow a smooth transition** from one emotion category (e.g., anger) to another (e.g., happy). However, to achieve natural human-like conversation, smooth rendering is crucial. *Mitigation:* 3D convolutional neural networks[24] will be implemented to extract transient features in the spatial and temporal domains and ensure smooth transitions.
  - The generated video is only in frontal view **without a natural head movement and minor expressions** when not speaking. However, to achieve a natural rendering, spontaneous head movements and minor facial expressions even during idle time are necessary. *Mitigation:* A large collection of human talking videos will be collected and modeled to learn natural head movements and minor facial expressions.
- Implementing **reinforcement learning** is crucial in the project, to exploit emotion to enable more efficient learning, better decision-making, and empathetic AIs. However, the number of iterations it generally takes a reinforcement learning system to achieve human-like performance is impractical in real-world scenarios. *Mitigation:* a natural selection simulator will be used to test how the AIs develop paramount secondary characteristics such as cooperation and competition. Research in this direction will start during Phase I but will be the main objective of Phase II.

Addressing these issues will represent a crucial step towards creating human-like responses in improvised interactions, for application on virtual AIs, humanoid robots or autonomous machines.

Intellectual Property

Emoshape's Emotion Processing Unit is protected by a **US patent #US10,424,318 "Method, System and Program Product for Perceiving and Computing Emotions"**, granted on September 24th, 2019 to the founder of the company, Patrick Levy-Rosenthal; a **PCT patent pending "Method of expressing emotions for AI systems based upon selected personality"**; as well as **trade secrets** for the **Emotion Synthesis, Emotion Reasoning** and **Emotion Processing Graph** machine learning algorithms.

The patents outline the details of the core technology in the EPU, along with how it is expected to fit within Emoshape's vision. The company will follow a similar patentability strategy to protect the innovations generated in this project, i.e., new methods for the emotion synthesis deepfake models.

NSF + I-Corps Lineage

Emoshape does not have any NSF lineage. However, this proposal has partial roots on NSF funding, as it builds upon research conducted by our partner, the University of Rochester, which was supported by NSF grant #1741472[25] "Audio-Visual Scene Understanding" – for designing computer algorithms that can understand scenes, with the aim of achieving human-like audio-visual scene understanding. The project team has not undergone any I-Corps training to date.

4. COMPANY/TEAM

Emoshape, Inc. Overview

Emoshape Inc. (NY), founded in 2014, is dedicated to providing an edge computing solution that teaches intelligent objects how to interact with humans to yield a favorable, positive result. We believe that the growing presence of AI, robotics and virtual reality in society dictates that meaningful emotional interaction is core to removing the barrier to widespread adoption and unlock applications in human-computer interaction that can have a significant societal and economic impact.

Emoshape's technology presents a new leap for AI on all fronts especially in the realm of self-driving cars, personal robotic, sentient virtual reality, affective toys, IoT, pervasive computing and other major consumer electronic devices. Applications in AI, Human-machine interaction, emotion speech synthesis, emotional awareness, machine emotional intimacy, AI's personalities, affective computing, Medical, Biometrics, Financial, Defense, Gaming, and Advertising will significantly benefit from the proposed Emotion Synthesis technology. We expect that our innovations enable intelligent machines to truly understand and sense what they say; sentient robots able to modulate their voice, facial expressions, and body language by the intensity of their emotions; advanced affective agents in soft toys; driverless cars capable of responding faster when making complex decisions by emotional reasoning; and robots developing their own unique personalities, learning from humans interactions. Our mission is to make the world a better place by giving intelligent machines empathy for humans, and create a positive future.

Emoshape innovations have been vastly recognized in the press, from internationally renowned media such as BBC, Forbes and Bloomberg TV, to specialized technology press such as the Neuro Robotics Magazine. There are over 30 articles[26] about Emoshape, and is also mentioned in 2 books[27,28]. For instance, Roberta Cozza, Research Director at Gartner, or Ray Kurzeill, Director of Engineering at Google, have recognized the major breakthrough of Emoshape's technology (see quotes in Figure 4).

Emoshape has raised $920K in funding to date from a VC and few LP's. Through pilot customers, such as Orange, Whirlpool, Siemens, or Volkswagen, among others, the company has generated revenues of $250K from prototype EPU microchip sales and cloud service contracts. The success of the present project will open the door to new market opportunities and Emotion Synthesis applications that are expected to drive significant company growth in the near future.

Key Personnel

Patrick Levy-Roshental, **Emoshape founder and CEO, Principal Investigator.** Patrick is a recognized expert, leader and entrepreneur in the IT and AI sectors. He has presented to the Artificial Intelligence council of the United Nations, won the IST Prize from the European Union, and has appeared in Forbes magazine. TEDx speaker, his 2006 worldwide-acclaimed invention, Virtual Lens, is used today by more than 1.3 billion people daily in Snapchat and Instagram. He moved to NYC to develop his passion and ideas surrounding bio-inspired emotion synthesis. He studied the relationship between cognition and emotion, and the influence of emotion on decision making, that lead to the underlying technology of the proposed effort. In particular, the EPU is based on Patrick's Psychobiotic Evolutionary Theory, extending Ekman's theory by using not only twelve primary emotions identified in the psycho-evolutionary theory but also pain/pleasure and frustration/satisfaction. Patrick is author of five patents, and has several publications on the importance of emotion to create intelligent machines. He is the precursor of several projects for applications of the Emotion Synthesis technology, and has conducted the preliminary research that the present project builds on. Based on his expertise, he has the skills and capabilities to lead and successfully execute the research activities of the envisioned R&D plan.

Dan He, Emoshape Data Scientist. Dan holds a PhD. in Computer Science by the University of California. He has worked as a research scientist on Algorithms, Machine Learning and Data Mining in IBM T.J. Watson, and as a senior research scientist, mainly working on AI+Healthcare, in which he participated and lead some of the core systems for IBM. He has published around 100 papers on top-tier machine learning conferences and journals, published many patents and won an IBM patent achievement award. Further, he has over 15 years of experience working as a consultant for many start-ups, having both research and development backgrounds. Dan he has worked with over 500 clients (as contractor, team lead, manager, CTO), including some Fortune 500 companies like Deloitte, Amazon etc., mainly on

the fields of NLP, Predictive Modeling, Recommendation Engine, and Deep Learning. Based on his expertise in machine learning and software development, he will support the PI throughout the project.

Partners

Emoshape will collaborate with the **University of Rochester** for the execution of this Phase I project. They will participate as a Subaward in this Phase I project. The research efforts will be conducted by:

Zhiyao Duan, PhD., is an Associate Professor in Electrical and Computer Engineering, Computer Science and Data Science at the University of Rochester, and is currently taking a sabbatical leave at Kuaishou AI Lab in Seattle. His research interest is in the broad area of computer audition, i.e., designing computational systems that can understand sounds, including music, speech, and environmental sounds. He is also interested in the connections between computer audition and computer vision, natural language processing, and augmented and virtual reality. He received a best paper award at the 2017 Sound and Music Computing (SMC) conference, a best paper nomination at the 2017 International Society for Music Information Retrieval (ISMIR) conference, a BIGDATA award and a CAREER award from the National Science Foundation (NSF). His research is funded by NSF, NIH, and University of Rochester internal awards on AR/VR and health analytics. He is one of the authors of the publication 'Speech-driven talking face generation from a single image and emotion condition'[29], that is one of the underlying research studies for the present project, and has collaborated with Patrick in the preliminary work (see Section 5) conducted so far to establish a proof-of-concept of the proposed technology. Moreover, he has also been the Co-Principal Investigator of the NSF Award #1741472 "Audio-Visual Scene Understanding", with the aim of designing computer algorithms that can understand scenes, i.e., to achieve human-like audio-visual scene understanding.

5. INTELLECTUAL MERITS

Emoshape's Innovation: technology description and background

What sets Emoshape's EPU apart is not only that it is the only technology capable of **real-time emotion synthesis**, but it accomplishes it using a **unique bio-inspired approach based on the human brain experience of emotions**. The EPU will infuse machines with the ability of understanding and generating any emotional state, and control in real-time their different facial micro expressions (FACS) and body languages. The EPU will also pave the way for the **next generation of NLP**, commonly referred to as Natural Language Generation. In this natural language processing system, the agent determines a thought or intent, and builds a meaningful response dynamically by putting different words together[30]. Therefore, the emotional state of the EPU can be considered to be a property that will get transferred to a Natural Language Generation system, leading to the generation of **Augmented Emotional Language** and emotion voice synthesis with WaveNet technology. These groundbreaking advances will represent a **step-forward in the human-computer interaction field, and in the pursue of sentient artificial life[31]**.

Emotion Synthesis: Wave computing and evolutionary theory

Emoshape's key innovation is its unique approach to high-performance machine emotion awareness, that makes possible to encode real-time unique emotional responses in artificial agents using Emoshape's EPU through Emotion Synthesis. The EPU technology is **inspired by computation neuroscience, evolutionary theory of emotion and psychobiotic evolutionary theory[32]**.

Emoshape is the only emotional technology that builds on the fact that human **emotions are not learned, but experienced**. The human brain is a remarkable pattern matching processor able to detect physical frequencies generated by external stimulus, and represent them in our consciousness as observable macroscopic oscillations (waves) in an electroencephalogram

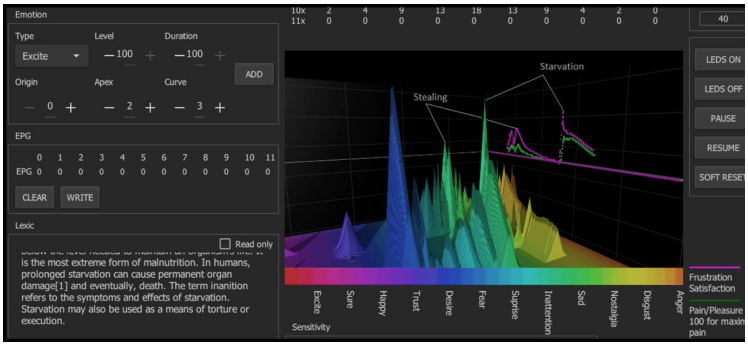

Figure 5. *AI connected to the EPU emotional response to the difficult question: "Would you prefer to starve or to steal?"*

EEG. Emoshape's EPU has been developed on that principle. The EPU enhances machine intelligence with emotional intelligence using **wave computing algorithms** where emotions are the result of frequencies representations. The Emoshape's EPU is capable of computing in real-time, the resonance, reverb, constructive and destructive interferences of multiple emotional waves' frequencies.

The EPU (Figure 5) is based on the primary emotions identified in **Ekman's evolutionary theory of emotion**[33]. The EPU algorithms effectively enable machines to respond to stimuli in line with one of the twelve primary emotions: anger, fear, sadness, disgust, indifference, regret, surprise, inattention, trust, confidence, desire, and joy. The EPU is also based on **Patrick Levy-Rosenthal's** (PI and company founder) Psychobiotic Evolutionary Theory extending the Ekman's theory by using not only twelve primary emotions, but also pain/pleasure and frustration/satisfaction levels, used to implement symbolic reinforcement learning. The Emotion recognition classifiers achieve more than 86% accuracy (ISEAR) on the conversation[34].

The real-time appraisal combined with an Emotional Profile Graph (EPG) computation functionality **allows the AI or robot to experience 64 trillion possible distinct emotional states each tenth of a second**. Emotional stimuli are stored within the memory bank with its associated cognitive and physical state. The EPG develops over time a bank of emotional associations and can communicate data to other AIs to achieve a realistic range of expressions and interactions designed specifically for each user.

A prototype of the EPU (Figure 6) has already been developed and tested. By analyzing external stimuli, including semantic appraisal, face emotion recognition, voice emotion recognition or other sensor data, the AI can virtually understand the user and engage in relevant conversations displaying sentient capabilities.

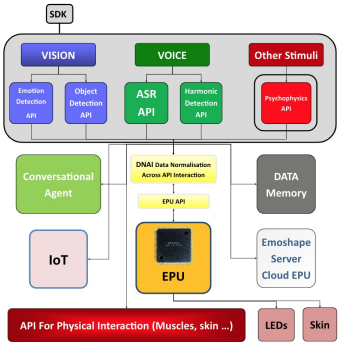

Figure 6. *EPU implementation*

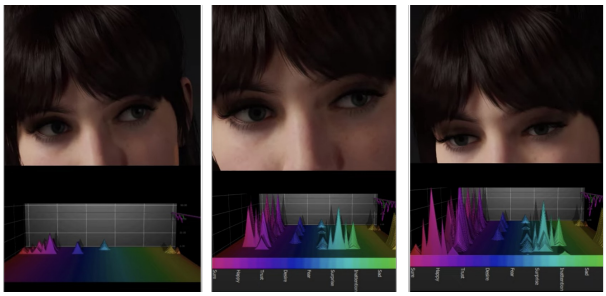

Figure 7. *Screenshoots from generated unscripted dialogue, displaying real-time emotional levels.*

The prototype has already been used to conduct preliminary proof-of-concept on autonomous metahumans, both to generate facial responses adjusted to emotional responses synthesized in the EPU, and to generate natural language improvised interactions. We have recorded live from Emoshape software development kit (SDK) a 100% unscripted dialogue with a metahuman (Figure 7, full video at: https://www.youtube.com/watch?v=Y6imP-MILgA). The metahuman displays real-time conversation and emotional appraisal, as well as voice emotionally controlled per word ESML, but still lacks full visual emotion expression. If we are able to calculate the corresponding facial emotional expression and connect the level of emotions in real-time to the FACS of the avatar, we will **achieve the first fully autonomous sentient metahuman**. This is the main goal of the work proposed in the present project.

Emotional Expression: DeepFake technology

In speech communication, emotion has a direct impact on the transmitted message[35]. However, predicting emotions purely from speech audio is quite difficult[36] and people heavily rely on visual cues in emotion interpretation[37]. To make emotion expression more realistic and improve communication, it is important for automatic talking face generation systems to render visual emotion expressions.

'Deepfake' has become an umbrella term for synthetic media applications which have their origins in Generative Adversarial Networks (GAN), such as StyleGANs[38,39], capable of creating computer generated human-like faces. Making deepfake videos involves the usage of a facial recognition algorithm and a deep learning computer network –a variational auto-encoder (VAE), which are trained to encode images into low-dimensional representations and then decode those representations back into images. A caveat to them thus far is how robotic and inhuman they appear. Using emotion synthesis technology in combination with deepfake technology could substantially improve their realism, through the understanding of a wide range of emotional states and the control in real-time of the different FACS, and could enable plausible use-cases in domains such as healthcare and gaming.

In preliminary work[xxv] by our collaborators from the University of Rochester, it was proposed the first neural network system that generates emotional talking faces from speech conditioned on categorical emotions, instead of inferring emotions from the input speech. The neural network (Figure 8) requires only a speech utterance, a reference face image, and a categorical emotion condition as inputs to generate a talking face video that is in sync with the input speech and expresses the input emotion.

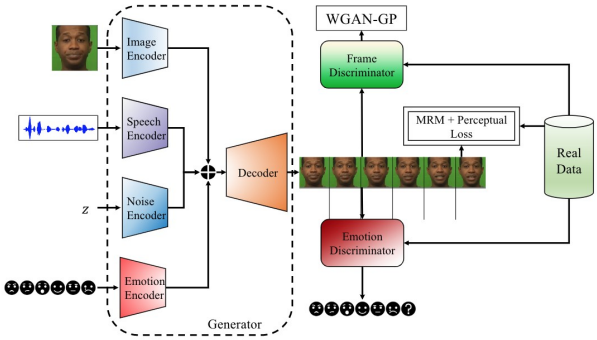

***Figure 8**. Overview of the proposed neural network system.*

During training, besides the mouth region mask (MRM) reconstruction loss and perceptual loss, the network employs two discriminative losses: the frame discriminator, which aims to improve the image quality of the generated video; and an emotion discriminator, which is essentially a video-based emotion classifier, with the inclusion of an additional class for fake videos.

Figure 9 shows examples generated from categorical emotion inputs; we will link the EPU with the talking face generation system during this project. Preliminary work on Emotion Synthesis DeepFake for the Emoshape-University of Rochester collaboration can be found at: [https://youtu.be/JBOHC_9II3o]. Although the results demonstrated that the generated talking faces were in sync with the audio and input categorical emotions, they also showed deficiencies for the development of realistic, human-like AI emotional expression (Figure 9).

***Figure 9.** Two examples comparing ground-truth (GT), baseline (BL) and proposed (OURS) videos for objective evaluation. Every 5th frame is shown.*

In particular, as a result of the original architecture, the system was only able to represent categorical emotions and was limited to short video generation without losing time coherence. Such limitations will be the main focus of the research proposed in the present Phase I.

In addition to talking face images, our collaborators have also developed multiple systems that can output talking face landmark points from input speech in noisy conditions[40,41]. These landmarks can then be used to drive off-the-shelf mocap-based face rendering engine to deliver deepfake videos in an alternative fashion. We will also investigate this approach as a side plan in this project.

Emotional Learning: Teaching empathy and self-preservation to AIs

A key aspect of Emoshape's technology is that it allows artificial agents to **develop a completely unique personality based on its user interactions and become more emotionally intelligent with each interaction by symbolic reinforcement learning**, hence no two will have the exact same personality. Using emotion machine-learning cloud computing and NLP, it is possible to force the AI emotional response to, for example, reproduce positive emotions when it has positive feedback from the user, i.e., the human is also experiencing happiness. This is implemented through rules on the pain/pleasure and frustration/satisfaction levels, that serve as intrinsic rewards for the AI. In this way, **Emoshape enables sentient and empathic AIs.** The approach is implemented using deep cognition capabilities through the Emoshape cloud service and is based on Patrick Levy-Rosenthal psychobiotic evolutionary theory[32].

Reinforcement learning techniques[42] have seen some of the biggest successes in machine learning so far, and are driving the advances in the field of autonomous vehicles. However, the number of iterations it generally takes a reinforcement learning system to achieve human-like performance is impractical in real-world scenarios. Infusing emotions (such as fear) with its appropriate choice-based reward/loss functions would substantially reduce the number of trials required to achieve a high level of accuracy. Research conducted by Microsoft[xi] on virtual cars driving in a maze, demonstrated that AI agents that incorporated intrinsic emotional drive (i.e., rewarded for minimizing stress/fear) did better than typically trained agents (i.e., trial/error): they drove for a longer period before crashing into a wall, and they explored more territory. Moreover, they performed better on related visual-processing tasks, such as estimating depth in

a 3D image[43]. Similar approaches can be realized using the **EPU on a life simulator to explore how different emotional drives influence AI systems learning capabilities**, personality development, and decision-making (i.e., for self-preservation). Emoshape has developed an evolutionary game based on genetic algorithms that will serve the purpose of exploring how emotion can give an edge to reinforcement learning and autonomous survival in various scenarios. Research on **emotion as a real-time computation of survival** will begin in Phase I, but it will be the main focus of Phase II.

Phase I R&D Plan

In this Phase I, we aim to design a system that can drive a full Emotion Synthesis DeepFake face with the emotion coded by Emoshape's EPU in real-time. The generated talking faces are expected to be in sync with the audio and express emotion states consistently with the EPU codes. Moreover, the proposed work aims to overcome high-risk technical challenges encountered in our preliminary work (detailed in Section 3), crucial to develop a realistic, human-like ES DeepFake. Namely, by the end of this project, our system will have the ability to reflect minor emotion expressions and smoothly transition between different emotions states. To establish the feasibility of our approach, Phase I will culminate on a proof-of-concept: our novel system will fully control a "metahuman" synthetic voice and face to demonstrate sentient capabilities through natural conversation. Initial efforts will also be directed to research emotional drives as a computation of autonomous survival and in reinforcement learning, crucial for some of the envisioned applications. The work plan for the present project is summarized in Table 2, below.

***Table 2**. Emoshape's Phase I work plan outline and timeline.*

WORK PLAN OUTLINE: Objectives, tasks and deliverables	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
Objective 1. Continuous-Valued Emotion State Expression												
Task 1.1. Dataset creation												
Task 1.2. Encoding continuous emotion states												
Task 1.3. Rendering continuous emotion states												
Objective 2. Time-Varying Emotion Rendering												
Task 2.1. Temporally coherent rendering												
Task 2.2. Natural transition of emotion												
Task 2.3. Natural head movement and micro facial expression												
Task 2.4. Talking face landmark outputs												
Objective 3. Demonstration of the ES DeepFake in a Metahuman												
Objective 4. Computation of Survival: Emotion and Reinforcement Learning												
Final Deliverables	1. Report describing the technical accomplishments and research outcomes of Phase I. 2. Prototype of ES DeepFake system demonstrating sentient capability, real-time, unscripted natural conversation and smooth facial expressions in sync with audio and emotional states.											

Objective 1. Continuous-Valued Emotion State Expression. Months 1-4

In Emoshape's ESML files extracted from text input, the emotion state representation is continuous-valued in emotion space (value of each state denotes emotion intensity). In preliminary research, we used categorical emotions in visual rendering. To make our model work with the ESML emotion states, we will create a new talking face dataset with continuous-valued emotion expressions to fine-tune our model.

- *Task 1.1. Dataset Creation:* Record an audiovisual dataset with annotated continuous emotion expression. Existing state-of-the-art datasets containing talking heads, such as MEAD[44] and CREMA-D[45], only use categorical emotions, hence are not suited for our needs. Actors will read text sentences with pre-extracted continuous-valued emotion states using the EPU engine. Such emotion states will be displayed as emotion curves to the actors, who will read the sentences following the prescribed emotions. Such curves will contain common emotion transitions (e.g., from calm to happy, from anger to depressed). The total duration of the recorded data should be around 10 hours. The text should cover diverse phonemes in English speech. We plan to hire about 100 actors to have sufficiently diverse facial and vocal expressions for training our emotion encoder.
- *Task 1.2. Encoding Continuous-Valued Emotion States:* Current emotion encoder will be redesigned to process continuous-valued inputs. In preliminary work, our emotion encoder used a two-layer Fully

Connected neural network to project the one-hot encoding for the six categorical emotions to emotion embeddings. Emotion in the EPU system is represented by an array of values (0 to 100), indicating the intensity of 12 emotional states. We will design a 2D Convolutional Neural Network (CNN) with multiple layers to deal with the increased complexity of the input emotion code. This CNN will be extended to a hierarchical recurrent neural network (HRNN)[46] in Task 2 to model temporal variations. Task 1.2 will focus on continuous state modeling. Training utterances will be segmented into emotionally coherent segments to prevent the encoder from being confused by the emotion transitions in the data.

- *Task 1.3. Rendering Continuous-Valued Emotion States:* The generated talking faces need to express continuous-valued emotions consistent with the emotion state input, rendering not only categorical emotions but also minor emotion expressions. Our preliminary work used a visual emotion discriminator that only focused on categorical emotions, missing the spectrum of minor expressions. We will re-design the emotional talking face rendering engine using instead a regression model as the emotion discriminator during training. This discriminator outputs the distance between the ground-truth emotion curve and the recognized emotion curve, and the training process will drive the talking face renderer to output video frames that carry continuous emotion states.

Deliverables/Milestones & Success Metrics:

1) Annotated audiovisual dataset of duration of 10 hours with sufficient voice and facial expression variation to provide rich training data for the engine; not only clear emotion states are prescribed and annotated, but vague periods such as emotion transitions will also be annotated.
2) Re-designed emotion encoder tailored for continuous valued inputs. To evaluate this design, we will take the encoder output and train a decoder to reconstruct the emotion input vector sequence and evaluate reconstruction errors. We will compare the proposed encoder with our preliminary design.
3) Re-designed rendering engine. Evaluation will consist of: (i) Objective evaluation: we will use a pre-trained face emotion regression model and assess the regression error against the ground-truth emotion curves; (2) Subjective evaluation: human subjects will be asked to draw arousal/valence emotion curves on a 2D plane in real time when watching the rendered video. Distances between the drawn curves and those mapped from the ground-truth emotion vectors will be calculated.

Objective 2. Time-Varying Emotion Rendering. Months 4-10

One limitation of our current method is that the generated video is not consistent if conditioned on long speech. Since in our preliminary work we used LSTM encoder in the generator and trained on a dataset with short spoken sentences, the artifacts get cumulated as the sequence gets longer. Another related limitation is that it does not allow a smooth natural transition from one emotion category (e.g., anger) to another (e.g., happy). For the talking faces to sustain natural, realistic conversations, both temporal coherence and smooth rendering natural transition are necessary features.

- *Task 2.1. Temporally Coherent Rendering:* By training with longer talking face sequences from the new dataset (Task 1.1.), the model will be able to capture richer temporal dynamics. We will also use hierarchical convolutional recurrent neural network (CRNN)[47] architectures at different time scales to encourage the model focus on both long-term and short-time dependencies. Furthermore, in the adversarial training part, we will add a sequence discriminator to distinguish between real and generated videos, focusing on temporal coherence.
- *Task 2.2. Natural Transition of Emotion:* To make sure that the transition from one emotion state to another is realistic, we will use unsegmented utterances that contain emotion transitions in each video to train the network. Furthermore, we will utilize the emotion transition labels to guide the network to pay special attention to the transition periods. On network architecture, we will use attention layers in the encoder and image renderer to generate natural transitions.
- *Task 2.3. Natural Head Movement and Micro Facial Expression:* To further improve the naturalness of the deepfake talking face rendering, we will explore the generation of natural head movements and micro facial expressions during speaking and idling time. To do so, we will track the 3D head pose in the training set using 3D body tracking techniques such as the SPIN model[48]. We will also infer natural head motion from the speech prosody such as the rhythmic head motion method[49]. We plan to investigate the correlation between head pose and speech utterances under the CRNN framework. When the speaker is not talking (idling time), the CRNN network will also generate spontaneous head

movements. Similarly, we will analyze micro facial expressions when talkers are not talking, and train the CRNN network to generate such spontaneous expressions.

- *Task 2.4. Talking Face Landmark Outputs:* As a side approach, we will also explore the possibility of outputting face landmarks in the CRNN framework for Tasks 2.1-2.3. This exploration will build on our preliminary work[35,36], and will be connected to an off-the-shelf mocap-based face rendering engine.

Deliverables/Milestones & Success Metrics:

1) Hierarchical architectures at different time scales. This will eliminate the inconsistency in the generated video when dealing with long speeches – minimum of a 2 minute video will be a measure of success.
2) Implementation of attention mechanism in the CRNN framework. This will help the generator to attend appropriate EPU and speech inputs for a more natural transition of emotion in the generated deepfake. The successful outcome will be viewed on a deepfake.
3) Integration of spontaneous head movement and micro facial expression generation. The successful outcome will be viewed on a deepfake.
4) Implementation of talking face landmark outputs in the CRNN framework as side output, and integration with an off-the-shelf mocap-based face rendering engine.

Objective 3. Demonstration of the ES DeepFake in a Metahuman. Month 11-12

The resulting model from Objectives 1 and 2 will be integrated in a "metahuman" from EPIC for the purpose of demonstration. The Proof-Of-Concept will demonstrate a sentient capability from a metahuman, enabling real-time natural conversation, hence the feasibility of the proposed technology.

Deliverables/Milestones & Success Metrics:

1) A report on proof-of-concept outcomes, including paramount findings, new technical challenges encountered, and areas for future improvement. The demonstration of the feasibility of the technology and the identification of next activities is the prime accomplishment sought after in this task.
2) A recorded unscripted conversation with the Meta-human, integrated with the newly developed face generating model and emotion encoder, that holds unscripted conversation for, at least, 3 minutes with synchronized voice modulation and facial expression.
3) Regarding 2), the primary face generation model will be the end-to-end approach described in Objective 2, but an alternative approach that renders face landmark points to drive a mocap-based face rendering engine will also be explored as described above.

Objective 4. Computation of survival: emotion and reinforcement learning. Months 7-12

A key success factor for sentient machines, will be exploiting emotion to enable more efficient learning, better decision-making, and empathetic AIs. Key aspects to study are whether emotion may give AI systems the crucial ability to generalize, and which emotional drives can improve autonomous survival of the AI without conflict of preservation of other biological lives (i.e. relevant for autonomous vehicles).

Emoshape will test this using an Evolutionary Artificial Life Simulation of Predator-Prey dynamics based on genetic algorithms. This natural selection simulator allows artificial agents and humans to interact with the world and each other. The goal is to survive long enough to reproduce and extend their survival solution through time. By changing their genome and mathematical brain, these agents can develop important secondary characteristics such as competition, adaptation and cooperation. Their EPU-enabled synthetic emotions and the player's emotion transmitted by multimodal input state can allow them to adapt and modulate the behavior of the agents in real-time, changing the course of the game.

Deliverables/Milestones & Success Metrics:

1) Report on the outcome of the Evolutionary Artificial Life Simulation, demonstrating adaption to attain a minimum level of survival, at least, doubling the time of survival obtained from the first test.

Critical Technical Milestones to Market

The successful completion of the present **Phase I will lay important groundwork towards commercialization of the proposed technology**. The proposed Phase I work will provide empiric foundation to: 1) affirm the proposed innovation is **technically feasible**, by proving that the combination of Emoshape's emotion encoder and talking face generation engine enables sentient AIs capable of real-time conversation much more natural and realistic than current AI chatbots; and 2) demonstrate the **added value of including emotional drives** to improve decision-making, and provide preliminary results of emotion as real-time computation of autonomous survival on the EPU through a life simulator.

During Phase II, the efforts will again be two-folded: 1) improvement of the ES DeepFake algorithms for expressing emotions without uncanny valley; and 2) further explore if **emotion can give an edge to deep learning and autonomous survival** and how to exploit it for faster learning or improved capabilities for more helpful AIs in in improvised scenarios for favorable, positive outcomes. **The results from this research will have immediate commercial applications** across several verticals and will pave the way for a new generation of sentient intelligent machines enabled by the resulting Emoshape software.

In the future, the same technology for face generation proposed in this project will be applied to humanoid robots. Although the emotion encoder and expression generator will work in a similar way, robots are complex, multi-disciplinary systems and achieving a human-like response is a tremendously difficult task. Mechatronic Engineering attempts at it have fallen short, only to capture a narrow range of possible expressions. Emoshape has already conducted preliminary work in this field in collaboration with Biomimic Studio and Artimus Robotics, who have worked on titles such as Star Wars, Prometheus and Westworld, to create a use case for a physical robot capable of facial and voice emotion synthesis: the emotion chip microcontroller capabilities will bring to life a sophisticated robot head by controlling 40 unique facial synthetic muscles. The developments of the present proposal will enable significant advances in this field, as challenges such as time coherence and smooth transitioning through continuous emotion spectra are also present in the pursue of emotion expression in robotics.

6. BROADER IMPACTS

The development of sentient, human-like, intelligent artificial agents can have a transformative effect able to permeate to all aspects of today society, and redefine the future of human-machine interaction.

The potential societal impact of its applications in the healthcare and caregiving industries, to name a few, is huge. Just in the USA, about 28% of older adults (13.8 million people) live alone, according to a report[50] by the U.S. Department of Health and Human Services[51]. Many of them are not socially isolated, but many of them feel lonely despite being surrounded by family and friends. Similarly, 2.2% of adults (5.4 million people) in the USA have Autistic Spectrum Disorder (ASD)[52], many with poor social skills and emotional difficulties. **Personal assistants with sentient capabilities,** augmented by vocal and anthropomorphic facial components, **can provide support to these elderly and impaired individuals, by providing a personalized aid with improved attention for these people,** as well as improving their tutoring, treatment and overall well-being. In parallel, this will reduce costs for caregiving and health organizations, improving efficiencies and patients outcomes. **Further, a sentient machine will be able to adapt its duties to the specific user,** enabling personalized approaches that will cater specific needs, hence further improving their service. This revolutionary technology can have the impact of improved personalized treatment, fostered trust and patient empathy, and an affordable and accessible therapy, it could potentially result in a healthier and happier human society.

Sentient machines can fulfill core psychological needs of humans – to be understood and communicated with at an emotional level, in some cases even better than other humans. Since humans are incredibly complex psychological beings, each one is limited by their own understanding of the world and operate within the framework of their worldview, and driven by personal goals, desires, and self-preservation instincts. An AI, by contrast, would have no need for any of those instincts and can be programmed to be a perfect emotional fit for a specific person or role – can have altruism built into them, along with other emotional functions that would elevate the experience of human beings interacting with them.

This project will represent a crucial step towards creating human-like responses in improvised interactions, whether it is for application on (i) virtual AIs, (ii) humanoid robots or (iii) autonomous machines. **Fast reacting empathic machines are needed in complex decision problems and, if biased towards human happiness and positive outcomes,** have an immense potential to impact numerous industries. For example sentient anthropomorphic cars have the potential of reducing car accidents; and tasks that currently require a human face and empathetic abilities can be delegated to robots, freeing human beings for more exciting and engaging activities, increasing overall productivity.

The possibilities extend way beyond the areas where AI is present today. Current AI reduces the world to universal statistical responses, but life is about diversity, its not a statistical prediction from the past, but a prediction on the future of it. Keeping the diversity of emotions will allow AIs to develop and respect diversity, as emotion is the currency of the experience in life, and AI is still missing that dimension.