**Improving elderly and special needs care through human activity recognition**

Dissertation Manuscript

Submitted to National University

School of Technology and Engineering

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OFPHILOSOPHY

by

Nate Bachmeier

San Diego, California

June 2023

**Abstract**

This manuscript presents a constructive design research project focused on utilizing computer vision to aid elderly and special needs patients. The research project encompasses the creation of various artifacts, including the analytics pipeline and the GraphQL interface, which enable the extraction and exploration of metadata from YouTube videos. These artifacts serve as valuable resources for investigating the potential of computer vision in assisting elderly and special needs patients. The proposed computer vision technique's effectiveness and efficiency demonstrate parsing videos using OpenCV, OpenPose, and custom Python scripts for experimentation and analysis. The results highlight this approach's utility in recognizing and understanding human activities depicted in the videos. This manuscript contributes to the growing body of computer science knowledge and its healthcare applications. The findings indicate that computer vision can play a vital role in enhancing the lives of elderly and special needs patients, offering a promising avenue for future research and development in this domain.

**Acknowledgements**

I would like to express my deepest gratitude and appreciation to my beloved wife, Pei-Chi, and our wonderful children, Eli, Maya, and Milo. Their unwavering support has been an invaluable source of strength and inspiration throughout my journey, and I am incredibly grateful for their presence in my life.

Pei-Chi, you are the rock upon which our family stands. Your unwavering love, encouragement, and understanding have been a constant source of comfort and motivation. You have always been there to lift me up when I faced challenges and to celebrate my successes with me. Your belief in me and your unwavering support have been instrumental in my achievements, and I am eternally grateful for the love and stability you bring to our family.

Eli, Maya, and Milo, my precious children, you bring immeasurable joy and fulfillment into my life. Your endless enthusiasm, curiosity, and unconditional love remind me every day of the beauty and wonder in the world. Your support and understanding during times when I needed to focus on my endeavors have been remarkable, and I am incredibly proud to be your parent. Your individuality and unique personalities continue to shape me as a person, and I am grateful for the lessons you teach me each day.

To my entire family, thank you for standing by me through thick and thin, for being my pillars of strength, and for believing in me when I doubted myself. Your unwavering support, patience, and understanding have fueled my determination to pursue my dreams. Your presence in my life makes every challenge more manageable and every success more meaningful.

I am forever grateful for the love, support, and sacrifices each of you has made along the way. Your belief in me has been the driving force behind my achievements, and I am honored to have you by my side. Together, we have created a loving and supportive environment that empowers me to pursue my passions fearlessly. Thank you for being the cornerstone of my life, and I look forward to many more incredible moments together.

Table of Contents

[Chapter 1: Introduction 1](#_Toc134428770)

[Statement of the Problem 2](#_Toc134428771)

[Purpose of the Study 3](#_Toc134428772)

[Introduction to Theoretical Framework 4](#_Toc134428773)

[Research Questions 6](#_Toc134428774)

[Significance of the Study 6](#_Toc134428775)

[Definition of Key Terms 7](#_Toc134428776)

[Summary 9](#_Toc134428777)

[Chapter 2: Literature Review 11](#_Toc134428778)

[Literature Search Strategies 11](#_Toc134428779)

[Theoretical Framework 13](#_Toc134428780)

[What is the role of data mining 17](#_Toc134428781)

[What exactly is artificial intelligence 20](#_Toc134428782)

[How does computer vision work 22](#_Toc134428783)

[What’s the role of Markov chains 24](#_Toc134428784)

[How are neural networks evolving 30](#_Toc134428785)

[How does intelligent agent modeling work 33](#_Toc134428786)

[How does neural network training work 38](#_Toc134428787)

[What is autoencoding? 47](#_Toc134428788)

[How does sequence analysis work 48](#_Toc134428789)

[How does recognizing human activities work 53](#_Toc134428790)

[Computer vision and autonomous driving 55](#_Toc134428791)

[How does the reproducibility crisis impact ML design 64](#_Toc134428792)

[Ethical Considerations of AI 70](#_Toc134428793)

[Summary 76](#_Toc134428794)

[Chapter 3: Research Method 78](#_Toc134428795)

[Research Methodology and Design 78](#_Toc134428796)

[Population and Sample 79](#_Toc134428797)

[Instrumentation 80](#_Toc134428798)

[Study Procedures 80](#_Toc134428799)

[Data Analysis 90](#_Toc134428800)

[Assumptions 94](#_Toc134428801)

[Limitations 95](#_Toc134428802)

[Delimitations 95](#_Toc134428803)

[Ethical Assurances 96](#_Toc134428804)

[Summary 98](#_Toc134428805)

[Chapter 4: Findings 100](#_Toc134428806)

[Validity and Reliability of the Data 100](#_Toc134428807)

[Results 101](#_Toc134428808)

[Evaluation of the Findings 114](#_Toc134428809)

[Summary 116](#_Toc134428810)

[Chapter 5: Implications, Recommendations, and Conclusions 118](#_Toc134428811)

[Implications 119](#_Toc134428812)

[Recommendations for Practice 122](#_Toc134428813)

[Recommendations for Future Research 122](#_Toc134428814)

[Conclusions 124](#_Toc134428815)

[References 126](#_Toc134428816)

**List of Tables**

[**Table 1** Design-science Guidelines (Hevner et al. 2004) 5](#_Toc134428817)

[**Table 2** *Survey search terms* 12](#_Toc134428818)

[**Table 3**  *Example Research Strategies for Classifying Movement in Video* 14](#_Toc134428819)

[**Table 4** *Activation Functions* 31](#_Toc134428820)

[**Table 5** *Principal Components* 34](#_Toc134428821)

[**Table 6** *Mean Time to Recover* 41](#_Toc134428822)

[**Table 7** *Example progressions of N.N. architecture complexity* 50](#_Toc134428823)

[**Table 8** *Ongoing Fees* 62](#_Toc134428824)

[**Table 9** *Threat Sources* 68](#_Toc134428825)

[**Table 10** *Alternative Research Approaches* 79](#_Toc134428826)

[**Table 11** *Download Process Infrastructure Services* 85](#_Toc134428827)

[**Table 12** *Processed Video Category Statistics* 102](#_Toc134428828)

**List of Figures**

[**Figure 1** *Should you purchase more stocks model* 24](#_Toc134429078)

[**Figure 2** *n-gram Examples* 26](#_Toc134429079)

[**Figure 3** *3 GANN Architecture (de Waal & du Toit, 2011, p. 399)* 27](#_Toc134429080)

[**Figure 4** *TensorFlow Architecture for MNIST Analysis* 28](#_Toc134429081)

[**Figure 5** *Multi-dimensional convergence (Kim & Cho, 2008, p. 1605)* 32](#_Toc134429082)

[**Figure 6** *BeeSmart Simulation (Wilensky, 2014)* 35](#_Toc134429083)

[**Figure 7** *Genetic Algorithm Process* 36](#_Toc134429084)

[**Figure 8** *Multi-Region Deployment* 43](#_Toc134429085)

[**Figure 9** *Paxos Consensus (Zhao, 2014, p. 196)* 44](#_Toc134429086)

[**Figure 10** *Durable Command Queue Pattern* 45](#_Toc134429087)

[**Figure 11** *On-device training architectures* 47](#_Toc134429088)

[**Figure 12** *Autoencoding architecture* 48](#_Toc134429089)

[**Figure 13** *NLP Analysis Procedure* 49](#_Toc134429090)

[**Figure 14** *Abstract Diagram of Differences* 51](#_Toc134429091)

[**Figure 15** *GAN Training Configuration* 53](#_Toc134429092)

[**Figure 16** *Network Structure* 55](#_Toc134429093)

[**Figure 17** *Taxonomy of Example Use-Cases* 56](#_Toc134429094)

[**Figure 18** *System Design* 57](#_Toc134429095)

[**Figure 19** *Training Configuration* 58](#_Toc134429096)

[**Figure 20** *Taxonomy of Participants and Example Challenges* 59](#_Toc134429097)

[**Figure 21** *Example Microservice Architecture* 60](#_Toc134429098)

[**Figure 22** *Preventative Maintenance System* 63](#_Toc134429099)

[**Figure 23** *Abstract pipeline* 81](#_Toc134429100)

[**Figure 24** *kinetic-700 video entry* 83](#_Toc134429101)

[**Figure 25** *Download Process Architecture* 84](#_Toc134429102)

[**Figure 26** *Video Processor Architecture* 86](#_Toc134429103)

[**Figure 27** *Video Processor Process Diagram* 87](#_Toc134429104)

[**Figure 28** *Skeletal Extractor Logic* 88](#_Toc134429105)

[**Figure 29** *Report.json schema* 89](#_Toc134429106)

[**Figure 30** *KeyFrame Selection* 90](#_Toc134429107)

[**Figure 31** *MovementTracker Logic* 91](#_Toc134429108)

[**Figure 32** *Tracking Persons* 92](#_Toc134429109)

[**Figure 33** *High-Level Analysis Process* 102](#_Toc134429110)

[**Figure 34** *Pose Output Format Body-25* 104](#_Toc134429111)

[**Figure 35** *Geo-distributing Traffic* 105](#_Toc134429112)

[**Figure 36** *Example GraphQL Query* 107](#_Toc134429113)

[**Figure 37** *GraphQL Response* 108](#_Toc134429114)

[**Figure 38** *Playing the cello (Video: -23ykna85DI)* 110](#_Toc134429115)

[**Figure 39** *Cello with label annotations* 111](#_Toc134429116)

[**Figure 40** *Playing hand-clapping games (Video: -MOpSXQ5ZcU)* 112](#_Toc134429117)

[**Figure 41** *Javelin Throwing Setup (Video: zVlBFLFkUNk)* 116](#_Toc134429118)

[**Figure 42** *Pouring beer (-5HJWCQ02Ds)* 121](#_Toc134429119)

[**Figure 43** *Clapping (0G1OirEz2OA)* 123](#_Toc134429120)

# Chapter 1: Introduction

A demographic change will significantly pressure the global healthcare system because people live longer, have fewer children and medical costs continue to increase (Piggott, 2016; Stone, 2017). When patients cannot afford the required care, the care quality decreases, or social programs must fund the difference. Demographic specialists predict that by 2050 nearly “80% of the global elderly population will be from low- to middle-income countries” (Mushsin et al., 2020, p. 1). Economic constraints within those countries will limit the effectiveness of their welfare programs and the availability of adequate services. Additionally, over one billion people globally have a limiting disability that requires additional support (Morris, 2008). Medical facilities need mechanisms to defuse the situation by reducing costs and deferring the transition to assisted living centers.

Inversely, the explosive growth across IoT, cloud, big data, and mobile (ICBM) continuously decreases costs and enables new opportunities. These technologies have the potential to revolutionize the healthcare and well-being industries. Academic and commercial vendors are continually delivering innovations across these domains. However, mainstream offerings primarily focus on measuring simple body metrics (Koreshoff et al., 2013). While these products provide incremental value, they do not move the needle. Nearly eight years later, the industry is driven myopically toward wearable IoT devices (Tun et al., 2021). Researchers concentrating on these areas make sense due to the low barrier to entry. Though, that same ease is commoditizing the product selection and stifling creativity.

Technology within special needs and elderly care settings has unique challenges and requirements (Ferati et al., 2016). These persons need unobtrusive systems that continuously monitor and respond to their behaviors. Specific vendors utilize voice-enabled Personal Digital Assistants (PDAs) (e.g., Amazon Alexa) to set reminders and record activities effectively (Tan et al., 2020). However, it becomes challenging to globalize these voice-specific technologies to assist non-native speakers and individuals with vocal disorders.

Assisted living facilities use trained nurses to mitigate these issues. Having a human inspect the patient visually is an effective but expensive tool. The median compensation rate for registered nurses is $75,330 annually ($36.22 per hour) (US Bureau of Labor Statistics, 2020). Due to the high cost, few patients have private nurses and receive fractional supervision. In contrast, video-centric monitoring and Human Activity Recognition (HAR) apply to a diverse population. When a person falls or drinks a glass of water, their skeleton moves in predictable ways. This consistency enables artificial intelligence & machine learning (AI/ML) to respond through cyber-physical systems (CPS). Businesses could deliver these capabilities economically and consistently across global markets, ultimately improving the quality of care at lower costs.

However, ethical concerns and privacy issues prevent researchers from collecting data at scale (Lei et al., 2021). Imagine the complexity that small-to-medium businesses face between vetting volunteers and ensuring diversity across participants. There are also budgetary considerations to deploying IP cameras and other CPS in numerous households (Shirazi & Shekhani, 2021). These challenges prevent quality research from occurring and improve patients’ quality of care. Instead, processes must exist to simulate these interactions and iterate toward more sophisticated systems.

## Statement of the Problem

The problem to be addressed in this study is the inability of elderly and special needs care organizations to capitalize on the effectiveness and efficiency of autonomous assistants (Blackhurn, 2021; Kim & Kim, 2021). Multiple industry-wide trends create the need for this technology. First, the number of practicing nurses has declined for several years (Kim & Kim, 2021). This labor shortage increases hiring and employee retention costs that the patients and welfare programs must cover. The funding gap is a global problem that does not impact all communities equally. For instance, rural special needs communities in South Africa have 57% fewer nursing visits than their urban neighbors (Besada, 2020). Newly industrialized economies like Taiwan, South Korea, Thailand, and Malaysia are experiencing challenges maintaining their long-term care programs due to growing costs (Phua, 2021). Domestic programs like Veterans Health Administration (VHA) and Medicare are not immune to these economic limits (Lei et al., 2021). Businesses and governments must control these costs and replace human labor with less expensive automation.

Beyond human and process issues are technical complexities in configuring prototype autonomous assistants. It requires multiple domain specializations like computer networking, embedded technologies, AI/ML, and distributed computing (Tun et al., 2021). Each cross-cutting concern adds complexity and reduces the probability that small teams can successfully provision their test environment. Furthermore, those difficulties limit other researchers from reproducing the results. These factors slow innovation and restrict the value researchers can contribute to the body of knowledge.

## Purpose of the Study

The purpose of this constructive research design is to study the effectiveness and efficiency of autonomous assistants detecting and responding to patient behaviors to reduce cost while improving consistency and quality for elderly and special needs care organizations. These organizations need human activity recognition models for numerous scenarios, such as handling patient falls (Shirai et al., 2021). Similarly, early dementia patients need monitoring capabilities to assist with discovering objects and providing task management (Lei et al., 2021). Collecting data from humans would be time-consuming, potentially dangerous, and rife with privacy concerns. The research design used in this study mitigates these challenges by using public video repositories such as YouTube. These libraries contain a diverse population performing labeled actions under varying physical characteristics such as weight, flexibility, and dexterity. Additionally, content moderators have painstakingly annotated the videos, enabling this study to focus on recognizing human behavior instead of bulk data labeling.

The study aims to create an extensible behavior classification model for indoor patient actions like exertion, insufficient nutrition, and episodic falling syndrome because of its medical importance and access to training data (Shirai et al., 2021). These situations negatively impact patients’ quality of life and may require additional medical resources. The study explores mechanisms for monitoring patients with a computer vision-based (CV) process that securely and reliably predicts patient behavior. This information could enable cyber-physical systems (CPS) to provide targeted care without requiring human nurses. That scenario would lower healthcare costs, increase patient access, and improve care quality.

## Introduction to Theoretical Framework

The design of experiments research creates purposeful artifacts and applies them to study a phenomenon (Hevner et al., 2004). Academic and business communities employ this method as a standard approach to Information Technology and Communication (IT&C) problems (Bryar & Carr, 2021; Peffers et al., 2007). It has well-defined guidelines (see Table 1) to implement a three-phased procedure. First, the researcher(s) must identify a domain-specific challenge. Next, that researcher creates artifacts that study this phenomenon. Third, those artifacts assess the topic and communicate answers to the research questions.

**Table 1**  
Design-science Guidelines (Hevner et al. 2004)

|  |  |
| --- | --- |
| Guideline | Description |
| Design as an Artifact | Design-science research must produce a viable artifact as a construct, a model, a method, or an instantiation. |
| Problem Relevance | Design-science research aims to develop technology-based solutions to important and relevant business problems. |
| Design Evaluation | A design artifact’s utility, quality, and efficacy must rigorously demonstrate well-executed evaluation methods. |
| Research Contributions | Effective design-science research must provide transparent and verifiable contributions to design artifacts, foundations, and/or design methodologies. |
| Research Rigor | Design-science research relies on rigorous methods to construct and evaluate the design artifact. |
| Design as a Search Process | The search for a compelling artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment. |
| Communication of Research | Design-science research must be presented effectively both to technology-oriented and management-oriented audiences. |

This study uses these guidelines and conceptual steps to identify a research-worthy topic and an actionable aspect. Next, it defines an abstract approach and implements a concrete proof-of-concept, the simulation process, to assess patient monitoring (via CV) and remediation (via CPS) technologies. Third, the artifacts expand the body of knowledge through the research questions. See Chapter 3: Research Method for more information.

## Research Questions

In alignment with the purpose of this study, and adopts the following research questions:

### RQ1

What is the effectiveness of autonomous assistance for classifying behaviors of elderly and special needs patients for care organizations?

### RQ2

What is the efficiency of autonomous assistance for classifying behaviors of elderly and special needs patients for care organizations?

## Significance of the Study

Human activity recognition (HAR) can improve elderly and special needs care by efficiently scaling out the visual coverage of medical facilities. Today, it is challenging to study HAR solutions within private residences. These issues stem from the system needing to record and share potentially privacy-sensitive situations, such as bathing or intimacy. Further complicating matters, the researchers must overcome the logistical challenges of finding representative samples, proving result reproducibility, and the economic overhead of multiple monitoring stations. Instead, this study proposes a research process using a physics simulator, animated actors, and virtual homes. The novel approach enables researchers to assess their CV algorithms across a repeatable configuration corpus. For instance, elderly patients falling is one of the most significant and avoidable reasons they need medical attention. This approach simulates this scenario, with each limb having distinct tensile strength, flexibility, and weight. When researchers can generate representative test cases economically, it unlocks the potential for faster product iterations and expands the knowledge body.

Cyber-physical systems (CPS) serve as a bridge between digital algorithms and the real world. These technologies need patterns and methodologies that react to intents discovered through HAR. Today, the fractured ecosystem spans multiple vendors, and it is cumbersome to assess holistic solutions. This research project aims to reduce this complexity with public video repositories. Further lowering the barrier to entry for study HAR within personal residences opens the door to future innovations not yet considered!

## Definition of Key Terms

The research study uses the following terms throughout the dissertation. This section includes a summary of their definitions. See *Chapter 2: Literature Review* for more information.

### Artificial Intelligence/Machine Learning (AI/ML)

Artificial intelligence is the design, implementation, and use of programs, machines, and systems exhibiting human intelligence. Its essential activities are knowledge representation, reasoning, and learning (Whitson, 2020). Practitioners use AI/ML processes to implement “fuzzy rules” that rely on statistical probabilities.

### Computer Vision (CV)

Computer (or machine) vision is a capability that extracts information from 2D and 3D images (Hornberg, 2017). CV processes can identify bounding boxes around objects, discover text, and classify facial behaviors, among other examples.

### Convolutional Neural Network (CNN)

A CNN is an artificial neural network in image recognition and processing domains (Nguyen et al., 2019). It mimics the biological structures of primate eyes by reducing image data through pooling actions. This process simplifies creating image masks for classification tasks.

### Cyber-Physical Systems (CPS)

Cyber-Physical Systems (CPS) are network-programmable devices that respond to digital messages through embedded capabilities (Aguida et al., 2020). It is a subset of an Internet of Things (IoT) domain. Practitioners use CPS to enable digital systems to interact with the physical world.

### Human activity recognition (HAR)

HAR processes identify human behaviors from motion feeds (Gorgulu & Tasdelen, 2020). It might assess several video frames and predict whether an actor is running, jumping, or sitting. These capabilities inform preventative and reactive controls to mitigate a situation. For instance, the process might detect a patient's fall and dispatch medical assistance.

### Internet of Things (IoT) device

The Internet of Things (IoT) attempts to widen the interconnectivity of computers by interconnecting objects (Commission of the European Communities, 2009). It monitors and reacts to the physical world using sensors, motors, cameras, and similar devices.

### Motion capture (MoCap)

Motion capture is a process that digitizes structural body movements for film and television production (Gan et al., 2020). Practitioners attach tracking sensors to a physical actor’s critical joints and inflection points. Computer programs record and encode those relative changes for virtual digital skeletal movements resulting in those agents exhibiting natural behaviors.

### Recurrent Neural Network (RNN)

An RNN is an artificial neural network for sequential data sets like natural language processing and time series (Boorugu & Ramesh, 2020). HAR processes can combine CV and RNN to observe and predict workflows. For instance, CV can detect an agent’s performing a high-exertion activity and use that output as input to RNN for forecasting that person will injure themselves.

## Summary

Healthcare costs are increasing, which creates the need for more automation. When patients cannot afford the required care, the quality decreases, or social programs must fund the difference. For many situations, like in-home monitoring of elderly and special needs patients, it is challenging to build that automation due to personal privacy and safety concerns. Researchers also encounter challenges spanning logistical, sufficient and diverse representation, and costs, among other entry barriers. After mitigating these issues, the research results are difficult and expensive to reproduce.

Implementing and verifying automation comes with a high barrier to entry, precisely due to personal privacy concerns, logistical complexity, ethical and cultural considerations, and procurement and configuration overhead. Researchers create frameworks to mitigate these privacy concerns (e.g., redaction), though these procedures are challenging. Beyond human and process issues are technical complexities in configuring prototype autonomous assistants. It requires multiple domain specializations like computer networking, embedded technologies, AI/ML, and distributed computing. Each cross-cutting concern adds complexity and reduces the probability that small teams can successfully provision their test environment. Furthermore, those difficulties limit other researchers from reproducing the results. These factors slow innovation and restrict the value researchers can contribute to the body of knowledge.

This study uses artificial intelligence and public video repositories to remove barriers to reproducing the experiment, personal privacy & safety restrictions, and logistical challenges. It implements these capabilities using open-source software and labeled training data from existing research. These recordings naturally incorporate people performing activities under differential physical configurations (e.g., weight and height). The study aims to demonstrate an extensible classification model that predicts the patient’s behavior from video sequences. While the study’s context is elderly and special needs, the application is more broadly applicable. For example, similar experiments could exist for monitoring childcare, performance coaching, house arrest inmates, and procedural audits. Regardless of the business context, researchers can solve critical cross-cutting concerns through HAR detection and CPS remediation.

# Chapter 2: Literature Review

The problem to be addressed in this study is the inability of elderly and special needs care organizations to capitalize on the effectiveness and efficiency of autonomous assistants through human activity recognition (Blackhurn, 2021; Kim & Kim, 2021). This research theorizes that computer vision (CV) can provide a consistent experience across a diverse global audience. Building autonomous assistants is challenging due to requiring multiple domain specializations like computer networking, embedded technologies, AI/ML, and distributed computing (Tun et al., 2021). Beyond technical constraints, potential privacy and safety from video monitoring create barriers to locating volunteer patients. Furthermore, those difficulties limit other researchers from reproducing the results. These factors slow innovation and restrict the value researchers can contribute to the body of knowledge.

The goal of this constructive research study is to provide an understanding of the effectiveness and efficiency of autonomous assistants using extensible human activity classification of video recordings. This study’s outcomes deliver this capability by extracting metadata from public video repositories and training a predictive model. Future research can leverage this information to modify the environment using programmable interfaces such as raising the alarm or applying other mitigations.

## Literature Search Strategies

This literature review used the Northcentral University Library (NCUL) to identify relevant peer-reviewed articles and books published from 2019 to 2022. It also includes foundational papers for historical context and generally accepted process standards outside this period. Students use NCUL’s Roadrunner search to aggregate results from industry-standard sources like the IEEE Xplore Digital Library, ACM Digital Library, and ProQuest.

A breath-first search scanned for surveys, challenges, and opportunities on the constructive research project’s core concepts (see Table 2). The breath-first search uncovered several themes that drove depth-first investigations. For instance, researchers are approaching hyper-scale ML training with custom hardware acceleration and continuous learning-at-the-edge methods (Plus Company Updates, 2021; Prapas et al., 2021). In other cases, themes like *Using Convolutional-Graph Neural Networks (C-GNN) for HAR* necessitate a sequential breadth-first search to contextualize supporting concepts. This search process continued until finding fifty unique documents. Next, bibliographical reviews for each publication extracted themes. Those sorted themes are available in the proceeding conceptual frame section, which presents each topic’s current state and direction from Table 2.

**Table 2***Survey search terms*

|  |  |
| --- | --- |
| Concept | Example search queries |
| Elderly and special needs industry state | * (elderly care or special needs) and industry * (global or internal) and (disabled or medical) |
| Computer vision (CV) | * computer vision or CV * computer vision and (surveys or opportunities) |
| Human Activity Recognition (HAR) | * (human activity recognition or HAR) and (computer vision or CV) * HAR (state-of-the-art or challenges) |
| Machine Learning (ML) Training | * (ML or machine learning) training and scale * distributed ML training |
| Physics simulation | * (Unity or ROS or robotic operating system) and (process or environment) simulation * (dynamic or synthetic or virtual) environment testing |

This chapter aims to frame the historical drivers and crucial decisions that shape the state-of-the-art for AI/ML and CV sciences. It approaches the problem starting with a low-level view of data mining and neural network technologies. Then it examines shortcomings across those areas driving deep neural networks (DNN) as the defacto solution.

## Theoretical Framework

A theoretical framework is a blueprint that communicates a natural progression of the phenomenon to be studied (Dickson et al., 2018). It is essential for quality research as it outlines a systematic structure of definitions, concepts, and relationships. Four core approaches exist for studying a business use case or phenomenon (see Table 3). This study’s blueprint derives from a constructive design science research (DSR) methodology.

DSR is one of the most common research methods for information systems and technology (Silvestrini & Sammito, 2012). These studies identify a problem, build artifacts, and communicate the implementation’s unique value (Hevner et al., 2004). In addition, many researchers follow this process to build proof-of-concept and execute case studies. This methodology is appropriate for examining elderly and special needs care solutions. After creating the system, it can support a targeted case study that measures its ability to deliver value.

**Table 3**   
*Example Research Strategies for Classifying Movement in Video*

|  |  |  |
| --- | --- | --- |
| Approach | Description | Study Example |
| Quantitative | Studies the magnitude of a phenomena | Measure the resources necessary to classify movement with embedded systems |
| Qualitative | Explores a concept without a numerical basis | Exploration of reasons movement classification fails |
| Mixed-Method | Combines exploration and studying the magnitude of these issues | What preparation steps reduce the costs of movement classification |
| Constructive | Produce artifacts to study a scenario | Create an algorithm for classifying movements |

### Fundamental Approach

Constructive research practitioners gravitate toward either Design Science Research (DSR) or the Constructive Research Approach (CRA). One of the critical differences between them is that DSR relies more heavily on existing theories, versus CRA does not explicitly require a base theory (Piirainen & Gonzalez, 2013). More recently, Iivari (2020) criticized the debate stating that constructive research must first and foremost produce high-quality artifacts. She advocates for “less theory, but better design theory (pg. 504),” especially within rapidly evolving industries like Information Technology and Communication. Zeller (2014) would agree with this position, adding success criteria that the artifacts are “challenging, elegant and useful.” This research project aligns with these requirements by connecting artifacts with patient needs and challenges.

### Central Concepts and Relationships

Here, the study presumes that CV and HAR can improve the livelihood of elderly and special needs patients. However, implementing those custom models is prohibitively expensive, and any research conclusions will be challenging to reproduce. This research project aims to mitigate these challenges by demonstrating CV and HAR methodology with simulated humanoids. It is beyond the scope of this dissertation to prove those methods are superior to existing and more laborious strategies.

There are multiple core concepts necessary to deliver this outcome. First, a literature review must examine the challenges and opportunities for elderly and special needs care. Quality research starts from customer challenges and works backward to find technological solutions (Bryar & Carr, 2021). In contrast, technology-first methodologies have a higher risk of not producing valuable outcomes.

Second, the central premise relies on exhibiting CV and HAR methodologies for predicting actions and behaviors. Computer vision (CV) processes extract data from image sources. Next, human activity recognition (HAR) processes must classify that data into distinct actions and behaviors, such as the person sitting or falling. Those requirements raise several questions within the literature review context. For example, what mechanisms are being built or deprecated? This research study does not aim to create a novel solution and plans to reuse existing methods.

Third, the research topic must train the ML model using a simulated environment with humanoid characters and virtual instruments. This situation raises implementation questions, such as trade-offs between industry-standard tooling, design patterns, and configuration nuances. The literature review must identify strategies that are likely to produce high-quality results. It is beyond the scope of this dissertation to implement proof-of-concept (POC) solutions for every potential combination.

Fourth, the artifacts must be high-quality and functional within a noisy environment. Meeting these expectations raises questions regarding ML training strategies. For example, do other researchers remove (or add) randomness to their DNN architectures? Are there specific situations that are more applicable for improving model quality? This constructive research study aims to incorporate these recommendations but stops short of directly comparing algorithms or methods.

Fifth, the literature review must uncover strategies for scaling the ML training and inference to the production scale. This sub-topic is crucial for bringing ML capabilities to public markets. However, it is also sufficiently complicated to populate multiple separate dissertations. Therefore, this dissertation only discusses literature trends for high-performance hardware, low-power hardware, and edge processing.

### Implementations and alternative framework

The proposed framework establishes capabilities that align with the business challenge of improving elderly care and special needs. It uses generalizable virtual camera instruments for CV and HAR experimentation with humanoid agents. Aspects of this framework appear in other publications (Banjarey et al., 2021; Gu et al., 2021). However, utilizing humanoid subjects with HAR and CV is not mainstream in surveyed literature.

An evaluation of alternative conceptual frameworks also took place. First, would an alternative virtual instrument be more appropriate? Instead, this study could predict HAR with accelerometers and gyroscopes (Gu et al., 2021). Wearable sensors have several advantages, such as following the patient worldwide. Researchers have also demonstrated applying these sensors to predict diverse action spaces (Nugroho et al., 2018). However, CV-based agents can extract more context from those same behaviors. For instance, a biosensor can predict that the patient is eating but not the food type. Instead of directly competing technologies, future solutions must integrate these heterogeneous sources.

An argument might exist that using humanoids is nonsensical and advocate for training the HAR models with public video repositories (e.g., YouTube). This approach has several benefits, such as realistic action depictions and freely available labeled data. However, training models on this real-world basis might be more challenging than the controlled and sterile simulation process. Furthermore, researchers can dynamically scale humanoid properties (e.g., weight) to assess model performance across parameter gradients. Future research could combine the frameworks with the video repositories validating the laboratory environment’s usefulness.

## What is the role of data mining

The four data-mining categories are association rule mining, clustering, classification, and regression modeling (Barua & Mondal, 2019). Association rules are patterns like if *X, then Y,* such as a person buying bread (X) is likely to purchase butter (Y). Clustering and classification are related strategies to group similar items into buckets. The critical difference is that classification knows the bucket labels beforehand (supervised), while clustering does not (unsupervised). For instance, a teacher gives their class a quiz and then maps them into groups by their assessment score (e.g., A, B) is a classification problem. Suppose they mapped the students on their favorite color. In that case, the groups are not deterministic, which is a clustering scenario. Regression modeling tries to find a mathematical equation that explains the observations. A classic example estimates housing prices using square footage, house age, and room count, among other features.

Across these high-level categories, numerous scenario-specific algorithms are available for different data sets. For instance, Apriori-based algorithms rely on the concept that subsets of frequent item sets must also be frequent item sets (Mejia et al., 2017). This property enables pruning the search space and timely report recommendations. Another use case comes from Self-Organizing Maps that cluster or categorize arbitrary data for anomaly detection (Sonmez et al., 2018). Then consider Ant Colony Optimization and Genetic Algorithms, which combine random guessing and regression modeling to iterate toward optimal solutions (Leios, 2017; Mirjaili et al., 2018). Other strategies exist to handle countless other challenges like dimension reduction (e.g., Principal Component Analysis) and brute force discovery (e.g., Parameter Sweeping) (Starmer, 2017).

### Organizational examples of data mining

Many financial investment firms rely on outcome-specific automated strategies to filter the sea of market data into a manageable number of options. For example, Fonskea and Liyange (2008) proposed a data mining strategy that tracks related companies' correlation (e.g., FedEx and UPS) and profits from deviations. In this case, both shipping companies will likely experience similar political and economic headwinds. Bhoopathi and Rama (2017) propose an Apriori-like algorithm to derive trading signals based on implicit associations between instruments (e.g., X and Y are inversely correlated). Hargreaves and Yi (2012) use a decision tree model to filter the Australian index on fundamental data (e.g., return on equity) from 2000 companies to a high-quality basket of the top six. Finally, George and Chang (2017) assessed the market interdependencies by transforming daily quotes into connected graphs.

### Explain challenges experienced using data mining

There is a joke that ‘70% of all statistics are made-up,’ inferring that the model is unlikely to work in practice without properly evaluating correlation versus causation. Carver (2007) explains that researchers focus on relevance, not “just seeing what we want to see.” Snee (2015) echoes that high-quality models are practical and explainable. Fonseka & Liyanage (2008) and George & Changat (2017) did not account for the contextually sensitive results of the Great Recession occurring in parallel. Bhoopathi and Rama’s association rules discovered tight relationships between Intuit (creator of TurboTax) and International Fragrance—with no economic justification. Aside from Hargreaves & Yi (2012), none of these approaches had a basis in modern market theory. For instance, correlations between price movements did not account for volume. The authors also limited their asset analysis to only primary assets instead of expanding into secondary assets. George & Changat (2017) determined that banks were the most critical aspect of their network but did not investigate interest rates, GDP, or consumer credit statistics. Bhoopathi & Rama (2017) could have transformed the data with a moving average to smooth out noise, decreasing false-positive rules.

### Enabling Machine Learning

Data mining enables transforming data into information. Researchers can build statistical models that predict outcomes when that information represents evidence for questions. Enabling this symbiosis requires well-defined objectives, or machine learning algorithms will lead to inaccurate solutions.

## What exactly is artificial intelligence

Dreams of artificial intelligence can trace back to philosophical debates in ancient Greece. Prometheus would mold handfuls of clay into images of the gods and later give life. Ideas sprouted from mathematics, biology, and computer science before eventually producing modern artificial intelligence. While these domains have unique perspectives, they collectively land in four categories of intelligent systems (Lukac et al., 2018). The first division asks if the system *thinks* or *acts*, or more precisely, can reason about the problem. These top-level categories contain subcategories of applications that mimic *humans* versus *rational* actors.

### Description of Technology

There are three high-level categories of artificial intelligence: rules and heuristics, machine learning, and deep understanding (Buchanan, 2005).

1. Before 1962, applications would rely on practical techniques for reducing the trial-and-error search space. This heuristic-centric approach is helpful for chess and other video game engines. Despite criticism for being naïve, many LOB (Line of Business) applications continue to leverage this technique successfully.
2. In 1963, Edward Feigenbaum and Julian Feldman’s *Computers and Thought* centralized many ideas across the computing industry. Their literature and new programming paradigms, such as McCarthy’s LISP, laid the foundation that became machine learning. Researchers use these tools to build statistical models that represent a situation. For instance, what else could you recommend if a customer purchases bread? Perhaps butter, jam, and deli meat.
3. In 1949, neural scientists found that the human brain transmits signals between a weighted graph of neurons (Lukac et al., 2018). Despite unlocking the biological key to mimicking cognitive learning, the processing power was unavailable until the early 2000s. Researchers use neural networks to extract patterns for nebulous problems that meet or exceed human capacities.

### Purpose and Function

Traditional software follows the model of *data* plus *rules* equals *outcomes.* In contrast, intelligent systems use data and outcomes to derive rules. This distinction can be valuable when the *rules* are fuzzy or not entirely understood. After extracting those rules into a model, researchers and engineering teams can predict actions across mechanical, thinking, and feeling tasks (Huang et al., 2019).

* ***Mechanical tasks*** are actions that are highly repetitive and benefit from automation. These are operations like turning on lights or assembly-line construction.
* ***Thinking tasks*** are operations that require analysis and rationalization. For instance, *Does this picture contain a hotdog*, or *Is this sentence grammatically correct*?
* ***Feeling tasks***, emulating interpersonal experiences, and expressing empathy toward the users. These autonomous systems might replace a call center or control support chatbots.

### Evolution of the problem

Numerous organizations begin their journey into intelligent systems with statistical modeling and variance analysis. These approaches work for many linear models but break down non-parametric functions (Waal & Toit, 2011). For example, a business wants to appraise houses given a collection of features about the home. Houses come in all shapes and sizes, making it challenging to compare those features directly. Instead, the appraiser must approximate a function considering these characteristics and their weighted importance. Meanwhile, another company must classify handwritten digits by mapping a 32x32 pixel image to its numeric value. Both scenarios and countless more require a mechanism to translate these non-parametric functions into parametric approximations.

### Nature’s solution

In biology, animal brains accomplish these tasks through meshes of neurons that transmit signals across connected synaptic (transforming) and activation (filtering) links (Keller et al., 2016). Later, that animal sees an object, and its brain encodes the image into a feature map. These features traverse the brain’s neural pathways and output a collection of responses, such as “the object is food and ten feet away.” Over time, the creature *learns* if those responses are correct and revise network weights to encourage or avoid similar situations. Data scientists and mathematicians replicate these ideas by calibrating edge weights, through backpropagation, on connected graphs called *neural networks*.

## How does computer vision work

Modern CV-based methods emulate primate biology across three distinct subsystems: neural dynamics, embodiment, and awareness (Ballard & Zhang, 2021). Researchers expand on these different subsystems to implement their specific use cases.

### Neural dynamics

Primates use retinotopy to map visual input from their retina to neurons. This process incorporates a random sampling and batching procedure to activate those neurons, with different combinations producing unique classifications. Marr (1982) proposed that machines could emulate this behavior to extract intrinsic images and functional constraints. His research shows that for every point within an image , it is possible to calculate its information level. Then, a smoothing function can remove the noise and produce object detection masks! Marr’s paradigm remains foundational to modern CV methodologies even forty years later.

### Embodiment

Bajcsy (1988) proposed that vision is an active process, and a hierarchy of decisions must occur. Each layer within the hierarchical map must encode the likelihood of a prediction within the context of the previous layer. For instance, when a person sees a cat, their brain uses different neuron groups to identify edges-to-shapes, shapes-to-labels, and aggregate labels to object names. Today, AI/ML practitioners call this construct the *hidden layers* within neural network architectures.

### Awareness

Around the mid-90s, researchers began exploring the notion of gaze control and fixation (Ballard & Zhang, 2021). They discovered that at least six separate systems stabilize objects within primate vision and implement a sophisticated data inventory system. For instance, when primates search for *a blue ball in the image*, their brains cache metadata to accelerate the gaze. Another critical service called fixation only persists memories with an associated high reward. For example, people safely drive to work without recognizing the preceding events because the routine operation did not produce new information.

In 1996, Kaelbling et al. proposed encoding these systems as policy maps that activate through an abstract reward function. Their notion of *reinforcement learning* explains how primates program their brain using visual information. Researchers have formalized this approach into a multi-process model where *reinforcement threads* combine to produce sophisticated composite decisions. Consider the problem of *Should I eat this food?* In this situation, parallel threads predict it is a *hotdog*, *hunger level*, and availability of *mustard*. Their aggregate response invokes an appropriate behavior based on visual information.

## What’s the role of Markov chains

A core challenge to applying basic statistics to real-world data is assuming that each action is independent. However, many scenarios contain a conditional state transition probability dependent on the current state. If the stock market falls 5%, should an investor buy? The binary question requires a contextually sensitive answer considering their net position (short the market), outlook (2008 financial crisis versus 2017 Trump bump), and similar factors. Markov chains provide the mathematical basis for making statistical models incorporating these dependencies (Kahn Academy, 2014). The hypothetical purchasing model (see Figure 1) begins with a state diagram representing the available actions. Then Monte Carlo solutions can approximate each edge’s weight by random sampling and recording the decisions. At the same time, multiple use cases can follow the same model, the scenario-specific decision weights. For instance, consider the differences between investing in (a) a 401k retirement account that only adds index funds versus (b) a delta-neutral (directionless) options trader. This trait is similar to other algorithms where efficient training requires relevant facts to specific questions.

**Figure 1**  
*Should you purchase more stocks model*

Diagram, schematic

Description automatically generated

### Markov Experiment

Many online tutorials recommend exploring Markov chains as a solution to predict the next token in a sequence. Mason (2020) maintains an open-source repository of Shakespeare plays, which is easy to mine for different related sentences. An experiment began with downloading each script and normalizing the text into a corpus of lowercase words. Next, an iterator constructs a word\_dictionary that maps n-gram tuples to a word bag to the immediately following values. Then traversal of the Markov model chooses a random starting point, then selects a random next word, iterating until a stop condition. Across the test iterations, tests of different n-gram sizes (degrees of freedom) ranged from one to six (see Figure 2). The higher the count, the more natural the sentences sound due to overfitting. Even at low n-gram terms, a frequent challenge arose from many unique words causing long sequences of static choices.

**Figure 2***n-gram Examples*

Graphical user interface, text, application

Description automatically generated

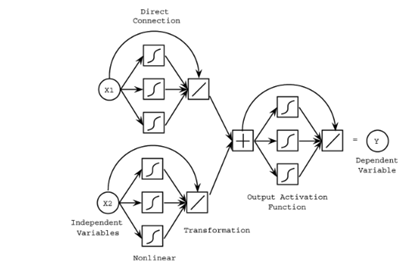
### Neural Networks

A Multi-Layer Perceptron (MLP) algorithm aims to map input features to a non-parametric function that approximates a set of outputs via an intermediary mapping function (the hidden layer). A fully connected graph can represent this structure. All inputs connect to the hidden layer, which connects to all outputs. Next, an iterative process forward-feeds examples through the network. Backpropagation updates the network weights and performs error corrections concerning the expected value (Ng, 2016).

According to Fridman (2017), backpropagation is a recursive process of taking the partial derivative of two logic gates and applying a weighted update. He expands on these connected graphs with an example of image classification passing through several layers (extracting edges, corners, object parts, and object identities). While the mathematical basis and engineering steps are somewhat procedural, the network architecture’s efficient design requires art and science.

Perhaps the artfulness comes from a lack of planning or awareness of how the *ensemble* of distinct training subsystems combines. There is no reason to assume that every node is fully connected or has an edge weight above zero (see Figure 3). A logical representation might consider feature ‘x1’ connected to N neurons that regress one output, with feature ‘x2’ implementing some classification pattern. These network segments produce collaborative signals to provide a more productive inference about the broader topology. These network segment microstructures remain present in more complex architectures. The solutions by both BellKor (2007) and Li et al. (2019) suggest that this assumption is generally accurate.

**Figure 3**  
*3 GANN Architecture (de Waal & du Toit, 2011, p. 399)*



### Neural Network Experiment

Consider the scenario of mapping 28x28 images of clothing to ten categorical labels (e.g., hats versus coats). The number of input features (neurons) is 784, and there will be ten output neurons—how many neurons should exist in the middle? Rosebrock (2019) provides an example solution (see Figure 4) to Fashion MNIST that begins with feature reduction through two max-pooling hidden layers and batch normalization. After cleaning, the solution uses a single 512-neuron hidden layer to predict one of ten output categories (with softmax). Reducing the hidden layer’s size to 128 or 256 has minimal impact on the cross-validation scores, though shallow values of 5 to 16 negatively impact accuracy. In this example, changing the activation functions (e.g., softmax to tan-h) creates more performance fluctuation than any other knob, with model accuracy ranging from 20 to 85%.

**Figure 4**  
*TensorFlow Architecture for MNIST Analysis*

Diagram, schematic

Description automatically generated

### Observations

The first and most critical step in any data mining exercise is determining the question and discovering supporting evidence. Until this action occurs, the business will unlikely have a successful deliverable and will spend excessive resources investigating irrelevant materials. After clearly articulating the business value, the engineer teams can perform broad filtration of data sources based on their ability to address those questions. During filtration, having a logical framework can improve the search process through partition pruning for the relevant data stores. For instance, if the business operates in Michigan, there is potentially minimal value in exploring Texas-specific data. After coalescing the supporting facts into a central location, cleaning and curation processes must confirm that the data is complete and pristine. Perfect information must be the right size and volume, or it might be incompatible with the analysis algorithms. For example, an instance learning algorithm expects individual records, not aggregate counts.

Markov Chains and Neural Networks are two strategies for making predictions on data through graph-like structures. Unlike basis statistics, Markov removes the need for independent actions and expresses them as weighted state machines. These state machines can improve workflow accuracy by guessing the next word in a sentence. Neural Networks and related MLP algorithms rely on weighted graphs and backpropagation to make predictions. While there is some artfulness, an alternative perspective asks if these are ensembles of small network segments. Evidence towards this interpretation exists in multiple advanced papers and helps to demystify the “machine learning black box.” It also means that several related concepts, patterns, and practices of data processing networks should also appear within more advanced neural network architectures.

## How are neural networks evolving

Frank Rosenblatt (1958) proposed the Mark I Perception as the first neural network architecture. This construct attempts to explain animals’ biological networks to *perceive* the world around them. The network consists of a collection of weighted sensors that converge into *one* learning circuit. Mechanical devices can replicate that implementation and train the connected weights to emulate animal intelligence with noisy data. Since this seminal paper, researchers have expanded deep neural networks to incorporate hundreds to thousands of connectivity layers. However, the specific configuration remains more art than science, with researchers manually iterating through trial-and-error experimentation (Ünal & Başçiftçi, 2021). The research field has two eras, with the first being the evolution of the artificial neural network (ANN) from 1989 to 2015. Then, the deep neural network (DNN) evolution became the primary focus from 2015 to the present. The industry is moving so fast that information beyond two years is becoming outdated.

### Artificial neural networks era

Perceptron was revolutionary, with its weighted signals triggering an activation function. This construct was insufficient for many scenarios and led to Multi-Layer Perceptron, which links a series of activation functions. Semantically, researchers can encode Boolean logic into these gates to derive more sophisticated insights. For instance, a network might contain two gates representing a person’s hunger level and food availability. Distinct signals can activate with each predicate to determine the overall scenario probability. That aggregate threshold can trigger an alarm or notification for the overarching decision to eat the food.

There are numerous activation functions, and a subset of the most common ones is available in Table 4. Originally researchers began with Sigmoid functions, which exponentially become a positive or negative one-value. However, this calculation is complex and slows down model convergence. A simple performance improvement came from using the tanh(x) function, similar to Sigmoid (Meta AI, n.a.). Now, researchers have chosen Rectified Linear Unit (ReLU) as the most preferred industry-standard algorithm (Ünal & Başçiftçi, 2021). Several scenario-specific variations like Leaky ReLU aim to scale and retain negative values versus truncating them entirely.

**Table 4***Activation Functions*

|  |  |  |
| --- | --- | --- |
| Activation Function | Formula | Description |
| Sigmoid |  | Mathematical function having an S-shaped curve with asymptotes at -1 and 1 |
| Tanh |  | A hyperbolic function that’s a ratio of sinh and cosh |
| ReLU |  | The most popular activation function |
| Leaky ReLU |  | An enhanced ReLU for incorporating scaled negative values |

### Architecture generalization challenge

Simple networks have poor learning abilities and are challenging to generalize to more sophisticated scenarios. Meanwhile, deep neural networks can learn intricate and subtle patterns but require more data before converging (Ünal & Başçiftçi, 2021). This trade-off causes many researchers to follow the principles of Occam’s Razor, which “promotes minimizing complexity and defending reductionism where possible (Oxford, 2022).” Calculating the most efficient and minimal network is an open problem, so researchers approximate with genetic algorithms. These algorithms aim to converge to a decent local optimum, not the global one. Genetic programming is an essential tool and recipient of significant scientific investment. Multiple dissertations could cover this topic, which is full of open problems.

Modern network architectures aim to simultaneously solve multiple objectives regarding weight and structural parameters to maximize fitness with minimal design (Ünal & Başçiftçi, 2021). Researchers can optimize various problem dimensions concurrently using ensemble methods, provided those subtasks have similar but not overlapping objectives (Kim & Cho, 2008). These subtasks typically mutate the network architecture through additive and pruning strategies until convergence, as illustrated in Figure 5.

**Figure 5**  
*Multi-dimensional convergence (Kim & Cho, 2008, p. 1605)*

Chart

Description automatically generated

### Deep learning era

Object detection and labeling tasks were among the first problems leveraging deep neural networks. Notably, in 2006, separate work by Hinton and Li led to the creation of ImageNet, a CV model for detecting twenty thousand labels based on fourteen million images (Ünal & Başçiftçi, 2021). In 2012, AlexNet incorporated graphic processing units (GPUs), reducing the error rate by 50% over previous CV architectures. Today, using GPUs over CPUs is table stakes and has opened the door to training across big data sets.

DNN architectures are particularly challenging to optimize because they contain high variability, multiple kernels, differing regularization scales, and untrained hyperparameters. Training hyperparameters control the model’s initial weights, learning rates, momentum factor, generalization, and the amount of training data (Jaisswal & Naik, 2021). These options influence several critical aspects of the final model, such as its sensitivity and degree of overfitting. Additionally, incorrect values can negatively impact training performance and defer model convergence.

Practitioners typically choose genetic programming or reinforcement learning (R.L.) for this procedure (Ünal & Başçiftçi, 2021). Data scientists can represent multiple expert systems as a connected mesh of R.L. models that search for ensemble methods, as Kim & Cho (2008) articulated. This mesh approach is standard for state-of-the-art architecture competitions like ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

## How does intelligent agent modeling work

Engineers consistently find that maintaining monolithic technologies requires substantial overhead. Alternatively, using microsystem architectures enables them to build and replace components rapidly in isolation. A similar idea exists with simulations that decompose the environment into multiple intelligent agents (see Table 5).

**Table 5***Principal Components*

|  |  |
| --- | --- |
| Aspect | Definition |
| Intelligence | The ability to reason about a problem |
| Simulation | An experiment that produces a statistical model |
| Environment | The universe contains the agents |
| Agent | An automaton that follows a predefined script |
| Objective | The goal of the agent |
| Tasks | The steps necessary to complete the objective |
| Notification | A collaborative or competing message between agents |
| Swarm | A group of agents |
| Choice | The random decision of an agent within its action space |
| Aggregate Choice | The net effect of multiple independent agent decisions |

A simulation experiment first identifies the environment, participants, and one or more objectives. Each participant, called an agent, attempts to complete their aim under guiding rules and principles. For instance, NetLogo’s BeeSmart environment contains multiple bees that try to maximize food production from various honey pots within a given scene (Wilensky, 2014). Initially, the swarm fumbles around until it discovers a food source. After some time, the colony will divide across multiple honey pots and compare site values with neighboring peers (see Figure 6). Eventually, the bees converge to the optimal configuration that provides the maximum food for the hive.

**Figure 6**  
*BeeSmart Simulation (Wilensky, 2014)*

A screenshot of a computer

Description automatically generated with medium confidence

While no individual agent (bee) understands the ideal distribution across the environment, the aggregate of independent decisions enables analysts to extract sophisticated observations about the broader objectives. It is also possible to quickly expand upon this simulation by designing expert agents, such as communication specialists, that propagate messages twice as fast. After defining the role and its local rules, the existing simulation can immediately incorporate those customizations.

### Genetic Algorithms (G.A.)

The Traveling Salesman is a classical graph puzzle that attempts to find the most efficient route through N-cities. Even with ubiquitous access to cloud computing, enumerating through an exhaustive search is not practical due to the combinations growing at (Keller et al., 2016). As the simulation continues to scale out, it requires a mechanism to prune that search space and quickly discover the optimal answers. The Theory of Evolution states that biology weeds out inferior strains through the Natural Selection Process (Darwin, 1859). Computers can replicate this model through Genetic Algorithms to converge on optimal configurations (see Figure 7).

**Figure 7**  
*Genetic Algorithm Process*

Diagram

Description automatically generated

The solution begins by modeling a potential answer as a vector of classification features. First, hundreds to thousands of randomly initialized instances run through the simulation to compute a per-instance score. Then a TOP-N ranking keeps the best instances, discarding the remainder. Next, a crossbreeding and mutation process mixes features from the fittest combinations to produce offspring. Those offspring cycle through this system thousands of times until only superior specimens remain.

### Multi-Level Agent-Based Modeling (ML-ABM)

After decomposing complex models into individual agents, a mechanism must aggregate the independent decisions into more macro observations. Cellular Automata (C.A.) paints this picture by grouping related swarms into “a hierarchical series of discrete systems (Makarenko & Osaulenko, 2018).” Through multiple levels of aggregation, agents can feed into swarms and those individual swarms into swarm networks.

For instance, a financial market environment has individual buy-and-sell participants who react to supply-and-demand fluctuations (see Figure 1). This example simulation contains thousands of personal portfolio accounts (agents) that frequently make rational transactions. An analysis could apply C.A. across these portfolios by aggregating many data points to improve the data’s usability for professional traders. However, an inefficiency exists within this design because some individual portfolios (agent states) are nearly identical. Like the risk-free rate, other aspects do not require the fidelity that swarms of agents produce. These situations can rely on ML-ABM to approximate irrelevant details (e.g., with caches) and enable fine-grained influence over critical decisions (e.g., with swarms of agents) (Hijorth et al., 2020).

## How does neural network training work

Model training aims to estimate the weights and connectivity structure for mapping a set of inputs to predict outputs. This process requires optimizing a cost function using a series of forward-feeding data operations followed by backpropagation (Fridman, 2020). Backpropagation is a mathematical procedure that compares the expected versus actual outputs. Next, it calculates each parameter’s partial derivatives and cascades updates to adjust the input weights accordingly (Lee & Yoo, 2021). Finally, after sufficient iterations, the network converges, which means that the difference between expected and actual outputs is within an acceptable margin of error.

### How are they scaling to millions of parameters

Recently, titans of the industry like Google Brain Team, OpenAI, and Uber Labs are scaling their evolutionary algorithms to millions of parameters without human intervention. Discovering the optimal architecture for CV and NLP problems at this scale is improbable due to its combinatorial nature and high costs per iteration (Lui et al., 2018). Instead, automation must represent the network as a traditional graph G consisting of nodes and weighted edges. Next, multi-level motifs augment the connectivity hierarchy to discover the impact of broad and narrow changes. For instance, these mutations might inject high-level filters or split the N-th hidden layer. In many ways, this is analogous to traditional fuzz testing, which exploits software through bit flipping.

Google DeepMind (2018) implements its search algorithm based on tournament selection. Each round begins by selecting the top 5% of the population and applying permutations from a discrete action space. After mutating the networks, only the superior offspring survive to the next round. This random search process is a genetic algorithm with an R.L. model steering its evolution to deliver the best of the breed through a simple search model.

### Modern scalability challenges

The sheer volume of matrix-based computations introduces challenges in scaling the training systems. Around 2012, researchers began using Graphical Processing Units (GPUs) over Central Processing Units (CPUs) to expedite model convergence significantly (Krizhevshy, Sutskever, & Hinton, 2012). According to some estimates, this hardware change leaped the state-of-the-art Deep Learning a decade (Ünal & Başçiftçi, 2021). However, researchers continue to require more extensive and sophisticated models. For example, GPT-3, an autoregressive language model, has over 175 billion parameters, a 10x increase over GPT-2 (Brown et al., 2020). More recently, Google Brain demonstrated an NLP translation model with over a trillion parameters (Fedus et al., 2022). This exponential parameter growth is likely to continue into the foreseeable future.

Several practical limitations exist to *scaling up* an individual server to support training a billion parameters. Instead, researchers turn to distributed systems patterns for horizontally *scaling out* across hundreds to thousands of commodity servers (Langer, He, Rahayu, & Xue, 2020). The orchestrator, parameter server, and worker node are three essential roles within these compute clusters. Worker nodes communicate with the orchestration process to determine the next training operation in the queue. Then, it fetches the shared parameter state and performs its work. Eventually, the operation completes, and the calculated gradients push to the parameter server. The parameter server is responsible for merging worker node updates into the shared state. This cycle repeats until the orchestrator determines that no more work is necessary.

### Fault-tolerant design requirements

The exponential growth in parameters and data volumes forces the training process to be carried out for days or weeks. That process must be highly reliable to withstand random errors during this period.

**Influence of Hierarchy.** Generally speaking, two mechanisms for modeling distributed systems are lists and trees. *Lists* can efficiently manage small groups of related nodes; however, they can become cumbersome with more massive sets. *Trees* allow for more expansive designs as the system can hierarchically describe the problem through multiple levels of control. Consider the difference between Domain Name Services (DNS, tree) and NetBIOS (list). NetBIOS can easily manage a small branch office because the simple flat list structure is *globalized*. In contrast, DNS has multiple subdomains, each owned by heterogeneous service providers. Since each subdomain holds a specific set of children, read and write operations can be *localized*.

**Influence of Partitioning**. Localized designs are inherently more performant and fault-tolerant because of the containment of both scale and blast radius (Vosshall, 2018). Imagine a scientific dataset that has grown to several petabytes in size. The storage network would need to decompose this logical file system into multiple blocks and replicate it across multiple physical servers. These physical servers will run into mechanical failures, such as disk corruption or power outages.

When these outages occur, other nodes must efficiently Set up, Challenge, and Repair (SCR) the missing data (Chen & Curtmola, 2017). The time necessary to perform that repair operation is proportional to the size of each block and the system’s ability to scale the reconstruction over multiple peers horizontally. Assume that 1TB of the dataset has entered a failed state and needs to recover across a 10GB/s network (see Table 6). If only one virtual peer has a copy of the data, the system will heal in 102.4 seconds. Then contrast that with the smaller block size of 128GB, which can economically be sprawled across many servers, reaching an MTTR of under a second.

**Table 6**   
*Mean Time to Recover*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Repair 1TB of Data | | |
| Block Size | Virtual Peers | Num Blocks | MTTR (seconds) |
| 1024 GB | 1 | 1024 | 102.4 |
| 8 | 1024 | 12.8 |
| 16 | 1024 | 6.4 |
| 512MB | 2 | 2048 | 51.2 |
| 16 | 2048 | 6.4 |
| 32 | 2048 | 3.2 |
| 256MB | 4 | 4096 | 25.6 |
| 32 | 4096 | 3.2 |
| 64 | 4096 | 1.6 |
| 128MB | 8 | 8192 | 12.8 |
| 64 | 8192 | 1.6 |
| 128 | 8192 | 0.8 |

**Influence of Fail-Over Groups.** Proxy servers and similar brokers operate on ephemeral requests and need fault tolerance from a different source. One strategy is maintaining a target group of service instances and monitoring their availability (see Figure 8). The monitoring can come from at least three reference points: (1) the network operating system, (2) the observed traffic of the broker itself, and (3) a local health agent on the service instance. The broker can use the Observed Health State Store (OHSS) to select the most appropriate receiver as new requests arrive. A recovery policy could also exist to manage any Service Level Objectives (SLO) of the backend application. For instance, if the backend application needs to be highly available, the broker could be augmented to trap specific exceptions and automatically route to another node. Other systems must optimize scenario-specific goals and metrics, such as more consistent response time, and choose completely different behaviors.

**Influence of Geo-Redundancy**. Cloud Service Providers (CSP) allow fault tolerance across multiple physical partitions (regions or availability zones). This design permits resiliency of entire data center failures without impacting uptime (see Figure 8). The scheme starts with deploying the service stack into two or more locations like Seattle and New York. Next, data store replication enables the sites to remain in sync. Finally, the user can discover the most performant service stack instance from a location-aware Canonical Naming Service (CNAME). That system can consider latency and other metrics, like the proposed Fail-Over Group solution.

**Figure 8**  
*Multi-Region Deployment*

A picture containing text, map, table, indoor

Description automatically generated*G*

**Influence of Consensus.** The physical distance between the sites forces the need for eventual consistency protocols that range in complexity from (a) the latest timestamp wins, (b) Paxos algorithms, and (c) Byzantine General’s solutions (Zhao, 2014). The latest timestamp wins are easy to understand. Still, they must rely on highly reliable distributed clock synchronization, an open research problem (Ting et al., 2014). Under Paxos (see Figure 9), multiple rounds of preparation, acceptance, and learning phases occur to gain consensus. This elegant protocol can efficiently reconcile a single systems image, provided none of the nodes are malicious. If malicious or erroneous nodes exist, 3f +1 cross-validations must occur (Zhao, 2014).

**Figure 9***Paxos Consensus (Zhao, 2014, p. 196)*

Diagram

Description automatically generated

**Influence of Protocol.** Message passing between components can either use reliable or unreliable communication. Unreliable handoffs can be helpful for best-effort or performance-critical systems, such as real-time video or sampled telemetry reporting. A reliable handoff is crucial for scenarios that mandate full and consistent accounting, such as user data or financial records. These fault tolerance decisions are not limited to the low-level transport protocol differences between User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). They also appear at higher application levels (see Figure 10). The actor can notify the Alice service directly; however, the message could become lost due to a network failure. Instead, they can place the payload into a command queue and remove it only after the server-side acknowledgment. When Alice accepts the event, she needs confirmation from Bob and Charlie before returning success. Bob stores the event in a durable command queue, whereas Charlie executes it directly. In either scenario, the client can reliably infer that handoff has occurred.

**Figure 10**  
*Durable Command Queue Pattern*

Diagram

Description automatically generated

### Scheduling-specific design requirements

The training process must also choose an asynchronous or synchronous scheduling architecture (Langer et al., 2020). Synchronous systems partition the training iteration (epoch) across the worker nodes and merge all responses before starting the next round. This approach has several advantages, such as simplicity regarding the design and maintenance. However, as the total worker nodes increase, so do the idle cycles from waiting on the slowest tasks to complete. Asynchronous systems aim to eliminate these idle cycles by prematurely starting the next training round. Since each worker node owns an isolated job, this technique is highly effective. However, the parameter server and orchestration processes must assume more complexity for this time optimization. For instance, what should the parameter server do when task outputs (gradients) are late or never arrive?

Langer et al. (2020) propose a collection of boundary conditions for asynchronous architectures. For example, the parameter server system should disregard any gradient exceeding K-epochs latency or change an individual parameter by X-threshold delta. They also advise using priority queuing over FIFO queuing for the incoming responses. Another solution is to use multiple parameter servers and periodically reconcile the differences between their local states.

### Continuous learning techniques

Countless researchers and product teams periodically retrain their models from scratch as they iteratively add more example data. This approach is intuitive to design but wasteful and could become impractical as model parameter counts continue exponentially growing. Instead, researchers are proposing training architectures that support *continuous learning*.

**On-device training architecture.** On-device training architectures aim to fine-tune generalized machine-learning models using local sensor data (Lee & Yoo, 2021; see Figure 11). This design begins with a cloud training process that consumes a public data set to produce a shared model. Next, a second training process will download and fine-tune the shared dataset to include personal data (e.g., a smartphone’s photo album). Lastly, an inference process uses the private model to deliver a personalized experience.

There are several essential advantages to this design pattern. First, the cloud-based training process does most calculations and has significantly more resources than an individual device. Next, the device-based training process can maintain data privacy and residency requirements because the personal data doesn’t need to be in the public data set. This characteristic could help overcome specific privacy concerns in regulated industries like healthcare. Third, the decoupled training processes refresh both models at a higher frequency.

**Figure 11**  
*On-device training architectures*

Diagram

Description automatically generated

**Stream-based training architectures.** This section is a placeholder for futureliterature review topics that may expand upon this point.

## What is autoencoding?

In its simplest form, “an autoencoder learns the representation or code by trying to copy the input to the output by encoding the input’s distribution into a low-dimensional vector (Atienza, 2018, p. 78).” Figure 12 contains an example architecture illustration with three features compressed into one parameter before expanding into three new features. The precise connectivity graph depends on the specific situation. For instance, researchers can use this process for scenarios such as colorizing images, denoising, replicating artistic styles, and intrusion detection, among other conditions.

**Figure 12***Autoencoding architecture*

Diagram

Description automatically generated

## How does sequence analysis work

Natural Language Processing (NLP) sits at the intersection of artificial intelligence, human language, and computer science.

### Language Parsing

NLP systems typically begin with sentence normalization, combining and annotating tokens, and performing custom business logic (see Figure 13) (Edureka, 2018). Using Lemmatiziation and Stemming strategies enables the parsers to reduce sentence variability, such as removing verb-tensing. Next, subsystems like Named Entity Recognition (NER) associate annotations with the words that discover the sentence’s critical components. After chunking related tokens together, the scenario-specific business logic can operate on a semantic text representation. Depending on the use case, these steps could be massive subsystems or single lines of code.

**Figure 13***NLP Analysis Procedure*

Diagram

Description automatically generated with medium confidence

### Deep Learning

NLP appears across various use cases like language translation, speech-to-text, and sentiment analysis. In biology, animal brains accomplish these tasks through meshes of neurons that transmit signals across connected synaptic (transforming) and activation (filtering) links (Keller et al., 2016). Computer scientists mimic this behavior with Deep Learning on Neural Networks, essentially weighted graphs. Generally, NLP architectures use Recurrent Neural Network (RNN) structures containing connectivity loops to previous layers (see Figure 14). More advanced designs include subnets for memory retention (see Table 7), encoding and decoding segments, and greater parallelization from attention vectors (Fu, 2019). Researchers and engineers can add or remove these subsystems to optimize a specific use case.

**Table 7**  
*Example progressions of N.N. architecture complexity*

|  |  |
| --- | --- |
| Algorithm | Description |
| seq2seq | Simple Recurrent N.N. (RNN) for a token sequence to sequence prediction. These systems are easy to implement but lack memory |
| Long-Term Short-Term Memory | Extends the seq2seq by including a “long term” cache to hold context information |
| Transformers | State-of-the-art solution for massively parallel NLP through attention vectors and position encoding |

**Figure 14***Abstract Diagram of Differences*

Diagram

Description automatically generated

### Feature Extraction Process

The first steps to any business intelligence problem are identifying the specific questions and locating facts to support answers (Snee, 2015). When researchers ignore this preparation, it produces garbage-in/garbage-out results. For instance, Alsudias et al. (2014) built an NLP system for predicting where the user was during the submission (e.g., a restaurant or nightclub). Their approach extracts keywords from Yelp reviews (using term frequency), business metadata (e.g., name and location), and tweet metadata (e.g., timestamp). These features flow into a random forest classifier that determines the user’s location with a 74% accuracy. However, using only the business metadata produces an 88% accuracy, indicating that these additional details provide negative value.

Researchers might have specific requirements to model social media users’ speech patterns and create new content in their voices. The Feature Extraction Process must therefore consider the user’s metadata (e.g., age and locale), the online community properties (e.g., forum name), the posted content, and any quality ratings (e.g., Facebook Likes). There are several considerations to augment this process. For instance, adding a filtration step to remove comments with negative ratings might create more well-liked personalities. However, it could also be advantageous to generate trolls that argue an alternative position, reinforcing the political situation that the other side is illegitimate or less sophisticated.

### Training Process

Around 2014, GAN (Generative Adversarial) Networks became the state-of-the-art approach to producing high-quality fabricated content (Fridman, 2020). These systems utilize a feedback loop between a Generative N.N. (GNN) and Discriminator N.N. (DNN). Each iteration outputs a ‘Deep-Fake’ asset and assesses its likelihood of being legitimate. This process enables both systems to learn from one another, continuously improving. According to Freidman (2020), detecting Deep Fakes is an arms race because advances in DNN naturally improve GNN results. NPAC leverages this methodology for self-teaching its systems to deliver more accurate content (see Figure 15). The organization’s solution uses the NLP transformer to improve parallelization over LSTM and a second RNN classification network. Periodic snapshots archive the content and model state for offline troubleshooting use cases during training.

**Figure 15***GAN Training Configuration*

Diagram, schematic

Description automatically generated

## How does recognizing human activities work

One critical CV application is to detect human activities from photos, images, and video streams. This capability is essential for personalizing systems across the healthcare, smart home, and safety industries (Gu et al., 2021). Adapting traditional ML tactics to human activity recognition (HAR) is laborious, error-prone, and challenging. Researchers mitigate these issues with deep learning models (Banjarey et al., 2021; Gu et al., 2021). There are several algorithm families used to model these predictions.

### Restricted Boltzmann Machine (RBM)

The first HAR implementations used Deep Belief Networks as their prediction basis. Training this solution was extraordinarily challenging and deprecated (Gu et al., 2021).

### Convolutional Neural Networks (CNN)

The preceding section examined the biological constructs enabling primates’ vision and the solution of nature. A neural network consists of three building blocks' input, hidden, and output layers. For instance, an animal image classification system might assign 64x64 pixel images into ten predetermined categories. This example requires an input layer with 4096 neurons, an output layer of ten neurons, and some hidden layers in the middle. Adding more hidden layers enables extracting more details from the image, similar to object edges (layer-1), ears (layer-2), cat’s ears (layer-3), and a tiger’s ears (layer-4) (Fridman, 2017). While more complex networks can extract more insights, it comes with the cost of needing exponentially more data to train the model.

Experts suggest a fully trained model requires at least ten observations per parameter (Snee, 2015). This situation can become too expensive and require model compression strategies (Cheng, Wang, Zhou, & Zhang, 2018). For instance, the input layer could feed into a series of pooling transforms that downgrade the resolution by averaging every 2x2 pixels. Another tactic might focus on connecting and evaluating local segments of neurons before outputting into global join constructs and prognostication output (see Figure 16). Meanwhile, other situations, like estimating housing prices perform better with fully connected shallow pipelines. While standard architectures exist for many classes of predictions, some experimentation is necessary.

**Figure 16***Network Structure*

Diagram

Description automatically generated

## Computer vision and autonomous driving

Researchers are applying CV to many health and safety systems like autonomous driving. Investigating these related use cases can uncover best practices and reusable patterns for elderly and special needs care.

Machine learning can enhance every aspect of the drive, from extending the physical parts’ lifespan to increasing the driver’s overall satisfaction. Figure 17 contains a non-exhaustive taxonomy of uses-cases for artificial intelligence in motor vehicles, such as reducing wear and tear and object detection. The central application topic is the primary deciding factor for item assignment within the tree. Many items, such as Voice Assistance (V.A.), could arguably live under a different pillar, safety. However, safety systems could use more traditional input interfaces in the same capacity, making this example fall under convenience.

**Figure 17**  
*Taxonomy of Example Use-Cases*

### Data collection process

Since covering each use case in full detail would fill multiple books, this section reviews these user scenarios’ commonalities. The lifeblood of these systems is data, and only through synthesizing information into knowledge can they be more adaptive. These processes require collecting telemetry, mining data, and modeling the interactions (see Figure 18). As simulations run across that model, statistical distributions form, leading to predictive capabilities. When the model’s complexity grows or the required accuracy increases, the learning system needs more examples to cover each scenario.

Consider the analogy of building an All-Wheel-Drive (AWD) feature that only knows about Florida’s flat tropical roads. Despite the engineers' best efforts, the vehicle will face challenges in Colorado’s ice-covered mountain climbs. Similar behaviors exist across the autonomous vehicle supply chain, demanding either (a) more data or (b) more erroneous assumptions. This trade-off introduces acceptable feature risks in specific situations (e.g., entertainment modules) and undesirable consequences for others (e.g., safety modules).

**Figure 18***System Design*

A picture containing text, map, indoor

Description automatically generated

Around 2014, GAN (Generative Adversarial Networks) became the state-of-the-art approach for constructing high-quality detectors and fabricated content (Fridman, 2020). These systems utilize a feedback loop between a Generative Neural Network (GNN) and Discriminator Neural Network (DNN). Each iteration outputs a ‘Deep-Fake’ asset and assesses its validity (see Figure 19). Under this process, both systems learn from one another, continuously improving their expertise.

**Figure 19**  
*Training Configuration*

Diagram

Description automatically generated

Automotive companies like Formula One use this methodology to synthesize more efficient race cars that can safely operate at high speeds (Smedley, 2019). First, a collection of features (e.g., car shape and weather data) load into the GNN and its simulation environment. Next, the output flows into a DNN, assessing the solution’s feasibility (e.g., wind drag and safety requirements). Then, the GNN modifies the solution’s parameters to search for a higher score (e.g., a faster car). After executing thousands of cycles, both networks converge with optimal solutions to deliver innovation (GNN) or detect problems (DNN).

### Safety Control Systems

Annually, 32,000 Americans die from automotive accidents, and another 2 million are injured (CDC, 2016). These statistics are unacceptably high and require innovations that increase all participants’ safety on the road (see Figure 20). Artificial intelligence can assist in these scenarios by collecting sensor data and predicting risks and opportunities. However, numerous open problems exist throughout the safety domain. These challenges should not discourage investments in these areas as they are essential to public safety.

**Figure 20**  
*Taxonomy of Participants and Example Challenges*

For example, manufacturers like Subaru and Lexus include audible collision alerts during lane changes or reversing. While these capabilities exist today, they are often incomplete models due to the high volume of edge cases, such as children fetching a ball from the street. Even after detecting the example child, several open problems span ethical and philosophical debates. Lex (2017) asksif avoiding the pedestrian requires killing the driver— what calculus dictates that autonomous decision? These situations might play out in fractions of a second, limiting the value of human intuition. Since concrete answers do not exist, machines must resort to static guardrails (e.g., slamming on the breaks or swerving) that could risk a multi-vehicle accident.

### Convenience Systems

Long monotonous segments are often necessary to arrive at the destination during a road trip. This requirement forces the driver to expel significant concentration relative to the mundane task. Instead, auto-pilot systems can take the wheel and allow the motorist to relax and participate in leisure activities (e.g., conversing with passengers). Some manufacturers tackle these needs with adaptive cruise control technologies. This approach is helpful in static environments (e.g., open highways) but encounters limitations in more dynamic environments (e.g., urban cities). Researchers are closing this gap by including more sensors that feed into sophisticated reinforcement learning algorithms (Fridman, 2020). However, continuing to scale these monolithic expert systems is challenging.

**Figure 21**  
*Example Microservice Architecture*

Diagram

Description automatically generated

V-TORCS (Virtual Open Racing Car Simulation) and other modern architectures address these issues using ensemble and multi-task learning methods (Li et al., 2019). Consider a decision process that feeds camera frames into an image classification Convolutional Neural Network (CNN) to extract objects and contextualize the environment (see Figure 21). CNN algorithms mimic an eye’s biological structures by normalizing neighboring pixel blocks to derive structure (Keller et al., 2016). These results flow into expert subsystems controlling the car, such as turning the wheel or accelerating. An ensemble of subsystem observations merges into a broader system-wide decision model that contains one or more I/O (Input/Output) controllers. Like other microsystem architectures, each subsystem’s implementation can evolve independently of peer components—enabling greater agility and innovation.

### Optimization Systems

Modern personal vehicles have a lifespan of over 200,000 miles and often travel 24,000 miles per year (Ford, 2012). Assuming a driver purchases a $25,000 car and keeps it that entire usable period, they will likely spend at least that much on fuel and repairs (see Table 8). Data scientists can improve this situation by introducing micro-optimization systems across the automobile. For instance, traditional cruise control maintains a specific speed (e.g., 70mph) without considering any environmental context. Meanwhile, a more intelligent system can factor in the road’s incline, the driver’s profile, and metrics about the trip to create a dynamic profile ranging from, e.g., 65-70mph. As this idea expands outward, it collects micro-optimizers that monitor all aspects of the driving experience, potentially saving hundreds of dollars in costs.

**Table 8**  
*Ongoing Fees*

|  |  |
| --- | --- |
| Line Item | Total |
| 200,000 miles @ 30 miles/gallon | 6,667 gallons |
| x $2.50/gallon | $16,700 |
| + Typical Repairs | $10,000 |
| Total Costs | $26,700 |

When the driver has advance notice that a component is likely to fail, they can schedule the maintenance and minimize costs. The Preventative Maintenance System (PMS) provides this capability by collecting component-level telemetry and looking for anomalous metrics (see Figure 22). For instance, an engine monitoring solution might observe the RPMs (Revolutions per Minute) and the electrical output. Sensor time series data are noisy and require a curation process (e.g., Kalman filter) to derive a stable moving average signal (Jackson & Rege, 2019). Next, the curated call flows into a Recurrent Neural Network (RNN), which uses sequences of previous tokens to predict future values (Keller et al., 2016). An anomaly exists when new observations deviate from these predictions and needs to surface in a decision-control process.

**Figure 22**  
*Preventative Maintenance System*

Diagram, schematic

Description automatically generated

### Smart City Integration

The future evolution of city planning makes urban areas highly connected with fast wireless networking and intelligent machines emitting enormous telemetry data volumes (Balduccini et al., 2018). The autonomous vehicle is a central participant in this ecosystem, as it transmits metrics to infrastructure (V2I) and other vehicles (V2V). These metadata feeds will enable more efficient traffic shaping, alertness to potential risks, and insights into driver patterns (Tong et al., 2019). For instance, each driver inputs their GPS destination and follows the route in an isolated silo. Tomorrow, aspects of those routing decisions can become centralized, resulting in less route congestion. While ideas of this ideal state are already coming to life in major cities like Seattle, Boston, and New York, the large-scale implementation is still years out (Cohen, 2013). Modernizing these areas will require significant infrastructure investments, consensus on V2X communication protocols, and machines implementing those standards. Machine learning technologies must synthesize those capabilities by making predictions by pairing vehicle-local sensors with ubiquitous cloud services.

### Observations

Three takeaways of this paper are (1) that artificial systems need to augment human processes; (2) a suitable starting place is safety and cost optimization; and (3) intelligent cities are several years away, requiring cloud services to fill that void.

The most powerful artificial intelligence applications use machines to enhance human capabilities rather than replace them. Motor vehicles contain many integration points. Machine learning can handle tedious aspects of the journey with greater precision and accuracy than humans. After freeing the driver from mundane work, she can focus on value-differentiating traveling qualities, such as talking with passengers and thinking about the day ahead. Meanwhile, intelligent systems remain vigilant in the background looking for risks like a child running into the road. Another collection of machine learning tasks revolves around cruise control capabilities. These systems keep us from harm’s way and reduce wear and tear on internal parts. Looking further into the future, integrating intelligent vehicles within smart cities promises more efficient traffic shaping and risk awareness. However, the necessary infrastructure investments will unlikely arise in the next decade. Instead, machine learning will pair vehicular telemetry with ubiquitous cloud computing to provide a similar experience.

## How does the reproducibility crisis impact ML design

There is an abundance of non-reproducible experiments because researchers do not account for nuances in the data collection (Rivera-Landos et al., 2021). These challenges originate from *Non-Determinism Introducing Factors* (NDIF), such as software updates and defects, hardware-specific differences, data uniqueness, randomization seeding, and dropout rates, among other reasons.

### Sources of Non-Determinism Introducing Factors

NDIF issues also originate from incorrect industry assumptions and unknown unknowns. For instance, software taint analysis is a process for discovering security-critical variables and parameters within an application. Generally, these tools parse the underlying bytecode through language-specific Reflection APIs. Next, they build call trees and use graph analysis methodologies to discover vulnerable code paths (Kilgallon et al., 2017). Researchers publish new techniques and present their findings using open-source products like FlowDroid, AmanDroid, and DroidSafe. However, the detection rate for those tools is highly dependent on configuration-specific settings (Qiu et al., 2018). This subtle dependency can be problematic for researchers and small businesses lacking the technical expertise to support multiple competing toolchains.

Recreating many foundational experiments is challenging and expensive, so researchers must assume that previous authors are correct. Unverified facts present a significant risk that can have cascading ramifications. Consider Majoranas, an extremely sensitive nanowire that operates at absolute zero temperatures and in extreme magnetic fields to measure individual electrons within quantum computers. This precision “plays an important role in protecting quantum information and enabling reliable computation (Langston, 2022).” Limited access to equipment forces many researchers to rely on computerized physics simulation processes. Recently, Pennsylvania State University could not reproduce several foundational studies and received opposite results (Frolov, 2021). After disproving basic expectations, researchers revisited other assumptions and found at least six false core tenants. This situation highlights the criticality of revalidating technological assumptions. The body of knowledge is never precise and continually evolves with new facts.

Negligence and maliciousness also create reproducibility challenges. Miyakawa has handled 180 manuscripts since early 2017 and identified 41 potentially fabricated experiments. When he requested raw for these studies, “more than 97% [(40 of 41)] pulled their publication request, suggesting a possibility that the data didn’t exist from the beginning (Miyakawa, 2020, p. 1).” The editor claims to spot these situations by looking for data too perfect, error rates improbable, and study impact too significant. Catching blatant lies is relatively easy though humans commonly only cheat a little (Ariely, 2009). Behavioral economists consistently demonstrate that people mispresent small details that lead to a better story with believably exaggerated results (e.g., a 75% accurate score becomes 82%).

### Influence of societal norms and ethical design

Hudson (2021) argues that researchers should focus on replicability over reproducibility. He identifies incorrect study design, not disproving the null hypothesis, wrong statistical methods, societal norms, and publication bias, among other factors. These factors impact research reproducibility. Therefore, researchers should accept that incorrect facts exist, and that’s because humans aim to prove what they believe. Douglas & Elliott (2022) responds to Hudson’s article, asserting it conflates value-ladenness with bias and mispresents values as evidential factors. They state that researchers are well-intentioned and aim to make reliable, repeatable studies. However, it is impractical for those practitioners to wait for results to be flawless, as this means science no longer evolves. Put another way, “All models are wrong, but some are useful (Denis, 2015, p. 3).” The quote infers that it can be challenging to include all aspects of the environment. These external factors can create deltas between expectations and reality. Academics should perceive the book of knowledge as continually developing principles and tenants, where more precise instruments will eventually supersede those ideas.

### Role of ethical design

Researchers need to manage ethical challenges that arise from their work. These issues originate from societal norms and internal biases. While several frameworks exist to guide the conversation, they can be ambiguous or focus on a subset of the problem.

Ethics are a system of moral principles that dictate the norms of a group. Societies implement these systems through social constructivism, enabling and constraining the group’s actions (Burr, 2015). Communities leverage this mechanism to assign truths and infer values about concepts (Gergen, 2010). Consider a project that seeks to prove that men are superior to women. Within a chauvinistic cohort, these results align with their worldview and are ethical. However, a diverse group would chastise the idea, regardless of methodology. Further complicating matters, moral identities are dynamic and evolve (or regress) over time.

Scholars need to understand their audience and the group’s customs. These social contracts limit the researcher’s influence and ability to solicit their work. These implicit rule sets vary between cohorts, making it impossible to remove these subtle biases entirely.

### Threats to validity

Four major categorical threats exist to making statistically accurate conclusions, leading to false, erroneous results (Parker, 1993). When designing high-quality experiments, the designers must be aware of these issues and their sources (see Table 9). Fundamentally, these challenges represent a degradation of the experiment’s confidentiality, integrity, and availability. These limitations prevent the generalization and reproducibility of research, resulting in the discrediting of publications and professional embarrassment (García-Pérez, 2012). Instead, researchers must decide on controls and procedures before collecting data.

**Table 9***Threat Sources*

|  |  |
| --- | --- |
| Source | Description |
| Internal Threat | Contamination by the research team |
| External Threat | Contamination outside of the study’s controls |
| Statistical Conclusion Validity | Results are arbitrary or non-reproducible |
| Construct Validity | Controls are not enforceable or consistent |

For example, if a participant needs to provide personally sensitive information, the data collection must convey trustworthiness (construct validity). Otherwise, the candidate will likely hold back data like side effects that are highly relevant to the research project. Without trust, the contributor might engage in activities that directly contradict the study (external threats). The research team might lack plans to deal with these confounding variables and deviate arbitrarily from excluding group members (internal threats). Since these results are now arbitrary, shoehorning outcomes into various statistical models until it lines up (statistical conclusion validity). An unlimited number of these permutations exist, and removing every situation is impossible (Parker, 1993). However, any procedure that reduces the influence of garbage-in/garbage-out experimentation is ideal.

### Internal biases

Researchers need to understand their internal biases. Everyone has historical and cultural defaults that lead to prejudices. These subtle classification differences influence language and construct our reality (Owen, 2017). Words matter, and one needs to choose them carefully. These biases sneak into our written and verbal communication, and they cause us to gloss over Diversity, Equity, and Inclusion (DEI) issues. For instance, the terms such as whitelist and blacklist have racial connotations. These modifiers become a sub-conscience reinforcement that one’s worldview is the only perspective.

### Sources of ethical frameworks

Numerous professional, regulatory, and advisory groups create frameworks that outline strategies for approaching ethical designs. These professional standards can contain conflicts of interest, hidden agendas, and inconsistent moral standards (Tan, 2021). The Belmont Report (1979) famously defines three core principles: respect for persons, beneficence, and justice. These tenants ask researchers to treat everyone fairly and avoid harm. However, even this simple statement has ambiguity.

After forty years, the ethical code requires modernization to align with evolving worldviews. Adashi et al. (2018, p. 1347) argue that the Belmont Report’s “distinction between research and practice is disappearing within the commercialization of present-day research.” Businesses actively debate the definition of “harm” and propose a notion of “harm versus setback.” This worldview states that any action that is not directly harmful is, at worse, an indirect setback. Roberts (2021, p. 15) proposes that researchers “must focus on risks of the research process itself, not outcome-related risks as downstream consequences are beyond the purview of ethical gatekeeping.” Facebook has a moral (and potentially legal) mandate to protect its user’s privacy. Under Robert’s definition, the social media juggernaut can ethically track relationships between billions of people. However, it is not bound to prevent malicious auxiliary use cases (e.g., election interference). While this position resonates with specific cohorts, it faces fierce opposition from others.

### Controversial Subjects

Many academic and business communities embrace Diversity, Equity, and Inclusion (DEI) concepts. These ideas are becoming mainstream, and that will cause them to become shared truths and social norms. Researchers that fight against this force are likely to find exclusion and isolation (Adashi et al., 2018; Owen, 2017). Instead, they should adopt social standards and assume “people are people.” However, this is often easier said than done. Human data sets contain numerous highly correlated variables (e.g., race and income). These statical properties prevent merely dropping an individual column and making the results racially neutral. Researchers can explicitly call out the risk in their findings, but fully addressing this situation is an open problem.

## Ethical Considerations of AI

Artificial intelligence is a scary black box that spreads malicious propaganda, destroys jobs, and seeks to destabilize honest citizens’ values. This statement is intentionally farcical, yet it also touches on fundamental concerns of ethical AI designs. People fear what they do not understand and use science fiction to fill these gaps. Within those futuristic worlds, machines become the dominant species that control every decision of an enslaved human population. However, several challenges prevent this transition of power from becoming a reality, such as intelligent systems that lack actual *intelligence* (Hole & Ahmad, 2019; Upchurch, 2018; Wildberger, 1996). Instead, organizations must assess these tools rationally, explore applications that enhance human capabilities, and remove undifferentiating overhead.

### Roles of Artificial Intelligence

Despite artificial intelligence already being well-entrenched in everyday life, there are concerns about its role. First, does the advancement of machine learning mean fewer jobs? Second, of those remaining jobs, are humans giving away control unnecessarily? Third, are those machines capable of manipulating the public to steal dominion?

**Role in Employment.** Before 1949, digging a ditch would take hours or even days with a crew of manual workers. After the invention of the backhoe, these jobs required less time with fewer employees. From the organization’s perspective, these efficiencies translate into faster time to market at lower costs. Meanwhile, the former diggers became displaced into new roles, repairing, operating, and supervising the machinery. Each of these positions requires entire supply chains of support. For instance, it takes factories to produce the backhoe parts, each staffed with hundreds of blue-collar jobs. Cities must also build universities and technical schools to train team members to fill these roles, expanding the job market.

Similarly, modern businesses actively seek methods to reduce costs and improve efficiencies through automation. The most powerful artificial intelligence applications use machines to enhance human capabilities rather than replace them (Boire, 2017; Heer, 2019). For instance, a person can write a more profound business case than a machine; however, the same device will have fewer grammatical errors. This dichotomy exists because humans specialize in contextualizing thought versus using patterns to make predictions (Scheler et al., 2019). Many professions exist as a combination of decision-making, pattern recognition, and mechanical tasks. Expert systems address specific job requirements; however, superseding the soft skills that unify these role components is challenging (Huang et al., 2019).

Specific low-skilled jobs, such as bank tellers and office clerical staff, are at risk of being replaced (Hamid et al., 2017). Similarly, expert pattern-matching tasks like identifying tumors in MRI (Magnetic Resonance Imaging) become commoditized through AI systems. Given the lower entry barrier, some low-skilled workers will transition to better-paying jobs operating sophisticated and commoditized systems. For instance, many workers cannot access foreign markets due to language and communication limitations. Artificial intelligence can aid these in these translation scenarios while leaving control with humans.

**Role in Decision Making.** Many decision-making processes can benefit from machines providing recommendations and validations. For instance, a court judge could use an intelligent system to assess how their sentencing aligns with existing norms. Perhaps the device predicts the defendant should receive five years of probation, while a judge considers fifteen years in prison. When the validation check expresses such a difference in opinions, it could suggest that unconscious bias is taking place and warrants additional considerations. That bias either provides ammunition for appeals processes or incarcerates people for unjustifiably long.

While this approach has much potential, there are concerns that professionals arbitrarily accept recommendations. However, these challenges occur everywhere automation controls the ‘last mile’ of decision-making. If the suggestion comes from a machine or peer, the person in charge of the process must be accountable for the final call. Blindly delegating control to machines is dangerous because learning algorithms are greedy, brittle, rigid, and opaque (Hole & Ahmad, 2019). Until artificial brains can rationalize abstract thought, humans must perform this task.

**Role in Manipulation.** Modern censorship does not restrict free speech; instead, it increases the noise and drowns the signal (Thomas, 2019). Fundamentally, marketing campaigns and propaganda machines follow the same process of Segmentation, Targeting, and Positioning (STP) (Kane, 2019). Delivering this objective requires pattern matching, content delivery, and human intuition. Automation is well-suited for these tasks and can use social media channels, like Facebook and Twitter, to connect with billions of people and manage significant portions of those interactions.

Congressional and media sources raise ethical questions about the ease of access to these capabilities for political manipulation. Unfortunately, these questions are mostly talking points rather than a call for action. Artificial intelligence has many abstract concepts that do not fit within the complex and opaque legal language (Guiffrida et al., 2018). For instance, machines cannot reason about their instructions, so how can the course hold AI/ML devices accountable? Perhaps the system designers should be responsible for their creations. However, the algorithms are primarily algebraic formulas controlled by end-users. Without a mechanism to define and enforce a standard operating behavior, it is impossible to expect a different outcome.

### Design Considerations

Two recent attempts to define this process for ensuring ethical AI are the European Union’s Ethics Guidelines for Trustworthy AI and the OECD’s Principals of Ethical AI (E.U., 2019; OECD, 2019). Both documents describe the need for artificially intelligent systems to be human-centric, transparent, explainable, robust, and secure.

### Human-Centric

Robotics’s Three Laws state that automation should not injure humans ignore people’s commands and protect their existence (Asimov, 1942). These rules lay a foundation for the idea that devices exist to cooperate and enhance humanity. Unfortunately, machines can’t reason and are bound to their program designs. Since machines cannot devise these criteria independently, it becomes the system engineers’ responsibility to enforce these requirements. Those decisions are predominately a matter of business priorities and vary across different use cases. For instance, a military weapons designer like Lockheed Martin views its human-centric role as protecting American interests at foreign nations’ expense. This perspective is radically different from other organizations yet equally valid.

### Transparent and Explainable

Artificial brains often rely on deep learning techniques through neural network solutions. These networks approximate a function that maps inputs and outputs through multiple non-parametric transforms. While data scientists can perform experiments to verify the model’s accuracy, they often cannot explain it (Gilpin et al., 2018). This limitation prevents broader adoption in places like the European Union, where the General Data Protection Regulation (GDPR) grants citizens a Right to Explanation.

Further complicating matters, neural networks learn the patterns we *ask*, not necessarily the ones we *mean*. For instance, Beauty.ai, an algorithm for rating female attractiveness, lost credibility due to only giving high scores to light-skinned candidates (Upchurch, 2018). This outcome was not intentionally malicious but the byproduct of not sufficiently representing minorities in the training set. Similar imbalanced issues occur across many real-world domains and require sophisticated data-handling strategies (Kaur et al., 2019). Even with expert data scientists, missing these edge cases and producing invalid predictions is possible.

### Robust and Secure

Engineers who become data scientists follow a different curriculum than their peers who become security specialists. This distinction in training is most evident in the lack of controls across artificially intelligent solutions (Lin et al., 2018; Sethi & Kantardzic, 2018). Malicious actors can influence these predicted decisions by inserting erroneous samples into the training set or directly attacking the probability distributions. For instance, researchers have shown that applying tiny amounts of distortion to images can change the graphic’s predicted class (e.g., cat versus dog) (Sethi & Kantardzic, 2018). If people cannot trust the classification algorithms’ integrity, how can mission-critical environments effectively use them?

### Observations

Artificial intelligence is a tool that can automate mechanical tasks, pattern-match data, and enhance human capabilities. Organizations can use these means to improve efficiency and reduce wastefulness. These innovations deprecate the need for specific skill sets and lower the entry barrier into other expert systems. While this causes an initial decrease in jobs necessary, entirely new industries follow shortly afterward. This promotion justifies the short-term pain when a society can replace low-paying jobs with high-paying alternatives.

Machine learning technology is too immature to delegate business-critical decisions. Instead, professionals should consider these technologies for initial recommendations and verify that their choices are free of unconscious biases. For example, a court judge should assess their sentencing aligning with a regression algorithm’s prediction and not blindly issue that verdict. Humans must maintain control of our actions and consequences. However, preventing machines from manipulating our free will can be challenging.

Laws cannot keep up with technology’s high-velocity innovation, causing businesses to define and self-regulate their ethical behavior. This moral desire must compete against existing business priorities without an official solution for maintaining accountability. Those priorities will vary significantly between organizations, as defining ‘human-centric systems’ is ambiguous. Moving past those challenges are issues with the fundamental integrity of neural network technologies. Implementing transparency and explainability are open research problems for all but the most trivial systems. After solving those issues, ensuring only inclusive training data use requires significant investments into unverifiable results.

These limitations bring the discussion around full circle to the beginning. Artificial intelligent systems are not ethical, evil, or corrupt. They are tools that automate everyday tasks and lower the barrier to entry. Users of that tool must know what these predictions mean and how they influence decisions. However, that is not the same thing as delegating control with impunity.

## Summary

This chapter contains the study’s literature review on modern AI/ML concepts. It began with foundational concepts like data mining and multi-level perceptron techniques that form the statistical basis for CV. These statistical models aim to emulate biological structures found in primates. However, primates have very sophisticated subsystems for embodiment and awareness that grant them contextually sensitive information beyond these simple models. Researchers use reinforcement learning to approximate Markov chains as decision policies. This technique has enabled machines to solve more complex use cases than ever.

Next, the deep learning era utilizes ubiquitous access to cloud resources and specialized hardware. Researchers can train models with nearly one trillion parameters to gain extreme predictions for challenging problems like natural language processing. They approach these state-of-the-art designs by assembling multiple reinforcement algorithms to mutate the model’s architecture. Like traditional generic algorithms, these expert systems crossbreed random model network connectivity until they discover the most efficient combinations.

There are several significant ramifications to this evolution. For instance, the computing resources necessary for training are growing exponentially, but the per-unit capacity is linearly increasing. This situation means that ML training must operate in highly distributed runtimes. Failures are likely to occur within these environments, and the orchestration system must account for error conditions. Consider the influence of discrete processors merging calculated gradients and the impact of network latencies. The processors began with relatively similar policy maps, mutated with an ensemble of RL updates, and must reconcile the changes. One could quickly fill out several dissertations on this topic alone.

Third, this chapter examines the reproducibility crisis and ethical considerations arising from these steps’ sheer complexity. AI/ML is a tool for enhancing productivity, and humans must remain aware that biases will creep into the design. These issues stem from systemic data correlations that are not always obvious (e.g., race and income). Controls and procedures must limit these factors and promote explainable AI.

# Chapter 3: Research Method

The problem to be addressed in this study is the inability of elderly and special needs care organizations to capitalize on the effectiveness and efficiency of autonomous assistants through human activity recognition (Blackhurn, 2021; Kim & Kim, 2021). This constructive research study provides an understanding of the effectiveness and efficiency of autonomous assistants for detecting patient behaviors for improving elderly and special needs care organizations. Like other projects, a high-quality research effort begins with a well-defined plan and stated outcomes. This chapter aims to meet these requirements by detailing the research methodology and its appropriateness. Next, it documents mechanisms for collecting data and analyzing that information. The chapter concludes by enumerating known assumptions, limitations, delimitations, and ethical assurances.

## Research Methodology and Design

Design science is a research methodology that creates and uses purposeful artifacts to study a phenomenon (Hevner et al., 2004). Academic and business communities employ this method as a standard approach to information technology and communication (IT&C) problems (Bryar & Carr, 2021; Peffers et al., 2007). The methodology comes with well-defined guidelines to implement a three-phased procedure. First, the researcher(s) must identify a domain-specific challenge. Next, that researcher creates artifacts that study this phenomenon. Third, those artifacts assess the topic and communicate answers to the research questions.

Many people erroneously believe one method is superior to another (Creswell, 2014; Jason & Glenwick, 2016; McCusker & Gunaydin, 2015). Instead, researchers must align the method with the research problem and purpose. Design science is appropriate for understanding the effectiveness and efficiency of autonomous assistants for creating an extensible human behavior classification model for elderly and special needs care organizations. The study considered and declined alternative quantitative, qualitative, and mixed methods. These approaches best align with problem and purpose statement variations (see Table 10). Suppose the objective is to compare treatment effectiveness or aggregate patient monitoring implementations. In that case, respectably, quantitative and qualitative methods are a better fit.

**Table 10**  
*Alternative Research Approaches*

|  |  |  |
| --- | --- | --- |
| Approach | Description | Example Use Case |
| Quantitative | Statistical modeling of a scenario | * Estimate the probability of an event * Stating a broad generalization * Cause and effect analysis |
| Qualitative | Non-numerical representation of a scenario | * Open-ended surveys * Exploration of needs * Investigating a local issue |
| Mixed-Method | Combination of both quantitative and qualitative | * Examining the breadth and depth of a topic * Exploring a scientific idea and then mapping it to use cases |

## Population and Sample

This study focuses on elderly and special needs patients’ behaviors, such as falls, malnutrition, and over-exertion. It derives these insights from DeepMind’s kinetic-700 data set containing 650,000 labeled YouTube videos (DeepMind, 2020). The video repository has a diverse population performing 700 specific tasks. For instance, 800 recordings are of people baking a cake and another 900 people brushing their teeth. This study aims to data mine this library from a breadth (many labels) and depth (label variation) perspective. Testing every video within the repository is economically impractical. Instead, a sampling procedure will select clips based on analysis complexity. For instance, kinetic-700 videos have poor lighting, blurry motion, and inconsistent reference points. While these characteristics are essential for understanding RQ2 (effectiveness), they might detract from the finite resources available to study RQ1 (efficacy).

## Instrumentation

The study has three aspects that require data collection: ML training performance, model accuracy, and inference performance. This information originates from the Amazon SageMaker services (AWS, 2021), which offer capabilities to build custom ML algorithms. Researchers essentially bundle custom automation and open-source tooling into a virtualized process. SageMaker uses public cloud resources like computing and storage to execute the experiment hundreds or thousands of times. It also integrates into TensorFlow 2 for collecting accuracy and performance metrics. These features reduce the complexity of building boilerplate instruments for many standard requirements.

These tools and services offer built-in consumable metrics using extensive developer tools such as Amazon CloudWatch and TensorBoard (AWS, 2021; TensorFlow, n.d.). The underlying services support custom metrics for investigating model efficiency. Implementing custom metrics beyond troubleshooting scenarios is beyond this project's scope.

## Study Procedures

This study aims to build a human activity classification model using the kinetic-700 public video data set. It implements an analytics pipeline for downloading videos, extracting metadata, and creating activity signatures (see Figure 23). A machine learning algorithm will process short video clips and predict agents’ intent based on their behavior. The algorithm models the subject’s skeleton movement changes into a sequence-to-classification model.

**Figure 23**  
*Abstract pipeline*

Diagram, schematic

Description automatically generated

This process begins with sampling the video stream from 24 frames per second (fps) to 6 fps. This operation aims to speed up model convergence and reduce operational costs. Next, similar to Das et al. (2019), the research will use the OpenPose library to identify the character’s joint positions in a given frame. Those matrix-encoded positions represent the input sequence to the model. According to the literature review, these matrices should be relative delta updates, not literal coordinates. Third, the matrices feed into an RNN-based algorithm (e.g., LTSM) for sequence analysis. This portion will use the standard Keras libraries for generating TensorFlow, a neural network. Lastly, a fully-connected layer extends the architecture to represent the classification action space and output. Specifically, the analysis performs the following high levels tasks:

1. Download the video from YouTube
2. Extract frames from the labeled segment
3. Extract skeletons from the video frames
4. Map skeletons across video frames into motion sequences
5. Perform object detection within the skeletal bounding boxes
6. Group the motion sequences into movement signatures

These constructs reside within an Amazon Web Services (AWS) account using cloud-first and open-source capabilities where appropriate. The deployment is fully automated using the AWS Cloud Development Kit (AWS CDK v2), which enables future researchers to reproduce the results efficiently.

### Data Collector Process

Content moderators for Google DeepMind built the kinetic-700 data set as a collection of YouTube video links and annotations for programmatic analysis. For instance, the following snippet describes the first video within the dataset as *clay pottery making* occurs during the 19th to 29th seconds (see Figure 24). Annotation segments are up to ten seconds long though the clip can be arbitrarily long.

**Figure 24***kinetic-700 video entry*

{

“---0dWlqevI”: {

“annotations”: {

“label”: “clay pottery making”,

“segment”: [

19,

29

]

},

“duration”: 10,

“subset”: “train”,

“url”: “https://www.youtube.com/watch?v=---0dWlqevI”

}

}

The data collector process is responsible for enumerating the kinetic-700 dataset and persisting it into local storage (see Figure 25). This process begins with enqueuing one message per video URL into a high-availability message queue with guaranteed *at least once delivery* semantics. Next, a containerized application dequeues the message and attempts to cache YouTube’s mp4 file into a network file system. Before contacting YouTube, the download process queries a NoSQL Key-Value store to confirm another container instance hasn’t already completed the video. After successfully caching the video, the download application deletes the message from the download task queue.

The cloud infrastructure supports scaling the download service container instances proportional to the queue depth. This characteristic also means the download service will shut down if additional messages aren’t waiting for processing.

**Figure 25**  
*Download Process Architecture*

Timeline

Description automatically generated with low confidence

This construct relies on several AWS services to orchestrate the download process and ensure high availability. The following table enumerates the essential components for managing this data flow.

**Table 11***Download Process Infrastructure Services*

|  |  |
| --- | --- |
| Service Name | Description |
| Amazon Simple Queue Service (Amazon SQS) | A secure, durable, and available hosted queue that lets one integrate and decouple distributed software systems and components (AWS, SQS, 2023) |
| AWS Fargate | A technology that you can use to run containers without having to manage servers or clusters of Amazon EC2 instances (AWS, Fargate, 2023) |
| Amazon DynamoDB | A fully managed NoSQL database service that provides fast and predictable performance to store and retrieve data (AWS, DynamoDB, 2023) |
| Amazon S3 | An object storage service that offers scalability, data availability, security, and performance (AWS, S3, 2023) |

### Video Processor

When videos arrive within the network file system (Amazon S3), an event triggers and enqueues into the OpenPose task queue (see Figure 26). OpenPose is a library from Carnegie Mellon that automates detecting human skeletons within 2D images using its proprietary models on GPU-enabled computers (Cao et al., 2019). This study packages the library into a containerized application that executes across a horizontally scalable GPU farm of Amazon EC2 p4gdn.xlarge instances (4 VCPU; 16GiB RAM; 1 NVIDIA A100 Tensor GPU). Like the downloader, the video processors monitor a message queue of pending tasks and avoid repeating work through a NoSQL status table. The awaiting message count influences the total video processor replicas, and the service will terminate when the queue is empty.

**Figure 26***Video Processor Architecture*

Diagram

Description automatically generated

Inside the video processor container is a simple message pump that pulls for OpenPose processing tasks (see Figure 27). The container fetches the associated video from the Video Store, extracts frames using OpenCV, processes them using OpenPose, and builds an analysis report.

**Figure 27**  
*Video Processor Process Diagram*

Diagram

Description automatically generated with medium confidence

The custom class *SkeletalExtractor* downloads the video file and uses the OpenCV *VideoCapture* class to access individual frames (see Figure 28). Next, the code iterates the annotation segment’s start-to-end offsets, retaining one frame every 500ms. Since kinetic-700 annotations are up to 10 seconds, a maximum of 20 frames are retrieved. The following code snippet outlines this process. Instance callers invoke the open method to bind the OpenCV VideoCapture class to the downloaded temporary file. Next, the caller generates the *Report* object by invoking the process\_frames method. This method is responsible for iterating through the video and collecting the sampled data. OpenCV offers APIs for fast-forwarding to specific offsets within the video, which minimizes the I/O requirements of the implementation. After selecting a frame, it’s passed to the OpenPose opWrapper construct for processing via the GPU-based hardware.

**Figure 28**  
*Skeletal Extractor Logic*

class SkeletonExtractor:

…

  def open(self):

    self.capture = cv2.VideoCapture(self.local\_file)

  def frames(self, step\_size\_sec=0.5):

    results = []

    offset = self.payload.start\_sec

    while offset < self.payload.end\_sec:

      self.capture.set(cv2.CAP\_PROP\_POS\_MSEC, int(offset \* MILLISEC\_PER\_SEC))

      \_, frame = self.capture.read()

      if frame is None:

        break

      results.append((frame,offset))

      offset += step\_size\_sec

    return results

OpenPose reports skeletons as *poseKeyPoints* 25x3 matrices that represent 25-body parts with respect to X-axis, Y-axis, and confidence score. The SkeletalExtractor consolidates these poseKeyPoints into the reports.json file as a list of *Frames* containing lists of *Bodies*. The report also references a frames.tar.gz file containing the input frames and skeletal annotated output images for troubleshooting purposes (see Figure 29).

**Figure 29**  
*Report.json schema*

{  
 'VideoId': string,

'TarFile': {  
 'Bucket': string,

'Key': 'frames/tarfiles/videoid/frames.tar.gz',  
 },

'Frames': [  
 {  
 'Offset': float,  
 'PeopleCount': int,  
 'Error': string,  
 'Bodies': [ [poseKeyPoints], ...],  
 }  
 ]  
}

### Frame Analyzer

When the video processor uploads the frames.tar.gz into the Frame Store (Amazon S3), the service raises an event triggering the Frame Analysis. This operation identifies and processes keyframes using Amazon Rekognition, a computer vision service (see Figure 30). Amazon Rekogniton can detect 25,000 objects, draw bounding boxes around people, and discover facial metadata (Mullennex & Bachmeier, 2023). The Frame Analyzer generates a second analysis report that overlays the OpenPose results with additional context. Suppose the skeletal movements suggest that *a person’s eating*. In that case, this overlay informs *what* they’re eating (eggs versus steak).

**Figure 30**  
*KeyFrame Selection*

class Report:

  def \_\_init\_\_(self, parent:ISkeletonManifest, bucket:str, object\_key:str) -> None:  
 ...

    response = s3.get\_object(

        Bucket=self.bucket,

        Key = self.object\_key)  
 ...

    self.object = loads(response['Body'].read())

  @property

  def key\_frames(self)->List[Frame]:

    return [

      x

      for x in self.frames

      if len(x.json['Bodies'])>0 and not x.has\_error]

## Data Analysis

The study procedure aims to process a significant subset of the kinetic-700 dataset and generate reports for skeletal movements with object detection overlays. This outcome connects to the foundational question of predicting human activity within video sequences for elderly and special needs care organizations. However, demonstrating that deliverable requires transforming the mountain of report.json definitions into *movement signatures*.

### Preprocessing the data

When OpenPose and Amazon Rekognition process frames, there are no assurances that the total people remain consistent, let alone the detected body parts. For instance, Alice and Bob might be dancing and twirling, which changes the perspective relative to the camera. The study must implement a *MovementTracker* process that normalizes the frame bodies into individual traces (see Figure 31). This operation requires cross-referencing bodies across the sampled frames and removing duplicate/erroneous information.

The business logic accomplishes this task through a two-step process. First, the extract\_people method returns each person and their associated metadata. This metadata includes information such as frame offsets and bounding boxes. Next, the find\_dups method analyzes the people list and returns duplicate records. These redundant people occur because of the parallel frame processing technique. Future research could optimize this approach to incorporate a tree structure and halt processing the same branches a second time. However, the goal of this study is not to write the most efficient code. Instead, this research project demonstrates human activity recognition and leverages duct tape with bread ties.

**Figure 31***MovementTracker Logic*

class MovementTracker:

  def \_\_init\_\_(self, report:Report) -> None:

...  
  def process\_report(self):

    people, metadata = self.extract\_people()

    duplicates = MovementTracker.find\_dups(people)

    unique\_people, unique\_meta =list(), list()

    for ix in range(0,len(people)):

      if ix not in duplicates:

        unique\_people.append(people[ix])

        unique\_meta.append(metadata[ix])

    return unique\_people, unique\_meta

The MovementTracker extracts people by iterating through each frame and body to recursively link the most likely bodies (see Figure 32). This operation predicts the best choice by evaluating the distance between two poseKeyPoints matrics and returning the sequence with at least three frames (1.5 seconds of video). While more effective strategies exist, this heuristic is sufficient for the study’s needs.

**Figure 32**  
*Tracking Persons*

def norm\_bodies(self, frame\_id:int):

    if frame\_id >= self.total\_frames():

       return None

    bodies = [

      np.array(MovementTracker.drop\_low\_conf(b))\* (1,1,0) / (self.image.size[0], self.image.size[1], 1)

      for b in self.report.json['Frames'][frame\_id]['Bodies']

    ]

    return bodies

@staticmethod

  def closest\_match(body, choices):

    best\_dist = 99999

    match = None

    if choices is None:

        return None

    for choice in choices:

        dist = np.linalg.norm(body-choice)

        #print(dist)

        if dist < best\_dist:

            best\_dist = dist

            match = choice

    return match

  def track\_person(self, frame\_id, body):

    sequence = [body]

    if frame\_id == self.total\_frames():

        return sequence

    choices = self.norm\_bodies(frame\_id+1)

    if choices is None:

       return sequence

    best = MovementTracker.closest\_match(body, choices)

    if best is None:

        return sequence

    sequence.extend(self.track\_person(frame\_id+1, best))

    return sequence

### Building the movement taxonomy

There are two phases to implementing an AI/ML process: training the model and operationalizing the capability. The analysis must confirm that these phases meet acceptable quality standards. Additionally, it must address the research questions from Chapter 1. This research project attempts to demonstrate extracting *intents* from dynamic and noisy video streams (see RQ1). A measure of the inference accuracy and the extent to which the scene contains noise must exist.

Tensor Flow 2 generates statistical information regarding the model training performance. These Key Performance Indicators (KPIs) describe gradient convergence, model accuracy, and various troubleshooting metrics. An analysis must confirm that the training configuration occurs efficiently. Suppose the performance is substandard. In that case, this research plans to investigate the defects and reconfigure the training service (e.g., Amazon SageMaker). It is beyond this study’s scope to create *a perfect model* and only seeks to demonstrate the concept. However, this study will validate that training is reliable and reproducible across positive and negative test cases.

## Assumptions

Researchers must be conscious of the internal and external factors influencing their studies. Making an assumptions inventory is essential to quality research because it communicates the implicit drivers in the design. There is an assumption in this research that the kinetic-700 files are compatible with industry-standard tooling. Suppose there are complications in ingesting and processing the content. In that case, the study can pivot to more depth over breadth when analyzing videos (e.g., fewer due to less automation). This approach is not as impressive but would complete the dissertation requirements.

The current plan also assumes access to a highly discounted rate for cloud computing resources. Amazon Web Services (AWS) has several programs for aiding researchers, like AWS Cloud Credit for Research and AWS Educate. Presently this study has funding through one or more of these programs and can pursue the entire project’s scope. Suppose that Amazon discontinued funding. In that case, the study would reduce the scale and focus on fewer test cases and humanoid configuration combinations.

This study makes several assumptions about the current industry state. It assumes that mainstream solutions, like Amazon SageMaker, deliver the capabilities necessary to build the core artifacts. This situation would allow the experiment to focus on the AI/ML components, not rewriting boilerplate infrastructure. Suppose the toolsets haven’t matured to an acceptable level. In that case, the study will simplify the training subsystem. Similarly, these can be simplifications for hosting ML inference endpoints if they are overly cumbersome.

The implicit assumption is that virtual agents can substitute humans in semantically similar configurations. It’s beyond this project’s scope to evaluate the validity of that assumption.

## Limitations

Limitations are internal and external factors that *implicitly* restrict the study from exploring all aspects of the problem. The previously stated assumptions also act as limitations. This subset would include technical feasibility, continued resource access, and real-world application.

This study aims to build a HAR classification model that supports an extensible set of activities due to the availability of example data. If the kinetic-700 isn’t sufficient, the study may need additional content from open-source repositories. These repositories could include YouTube, among other sites.

## Delimitations

Deliminiations are internal and external factors that *explicitly* restrict the study from exploring all aspects of the problem. This study utilizes video footage from third-party resources that might behave outside the desired human activities. These distinctions arise from external factors beyond the researcher’s control with respect to the underlying data. HAR is also a vast concept with virtually unlimited permutations.

## Ethical Assurances

Northcentral University’s Institutional Review Board (IRB) must issue a statement covering ethical concerns, privacy violations, or undue harm risks. This study does not involve human or animal participants; in this respect, it meets the IRB requirements by default. The research utilizes computer simulations, which doesn’t raise privacy or ethical concerns. This study uses open-source third-party video as the input, which mitigates ethical matters of personal privacy. Furthermore, the people within the video clips agreed to the YouTube terms of service and no longer have an assumption of privacy.

Medical facilities have a business requirement to collect private information from patients. While building a system that stores and retrieves this data is relatively trivial, several specific considerations influence the final implementation. Which users can issue queries against the datastore? What maintains the confidentiality of these records? How will auditing and compliance reporting work? Does this data have legal or regulatory implications? Answering these questions produces a model of acceptable risks and identifies business policies requiring cybersecurity enforcement. These enforcements protect the business against negligent and malicious attacks that could harm the integrity or reputation of the brand.

The principal objective of any business is to execute its mission in the most efficient manner possible. Delivering on that mission requires choosing between acceptable risks and desirable conveniences (Mickens, 2018; Dai Zovi, 2019). For instance, many small to midsized business owners lack the expertise to run a domain controller or email service. Employing dedicated staff retracts from resources that could provide value differentiation towards its core competencies. Contracting a consulting firm would be less expensive but lacks the deep economy of scale discounts from Microsoft Office365. While financial factors influence many decisions, the security and compliance teams must assess the privacy and availability risks. Not all decisions originate from the leadership and often from internal department requests. For instance, a data science team might require a Juypter Notebook server with access to a production database. While that team has enough knowledge to be dangerous and deploy an operational instance, they might lack a broader understanding of business continuity requirements (Brown, 2015). What physical host controls this instance? Does the database connection use encryption? How is backup and restore scenarios handled? Until understanding these subtle decisions, it is impossible to determine if a failed server hard drive will lose three minutes or years of productivity.

These decisions must influence the study’s data storage design to be secure, reliable, and durable. In this context, the *seed* data is not confidential and comes from public repositories. However, there are risks that the *result* data can become corrupted or destroyed. That situation would risk the dissertation process completing on time. This constructive research project mitigates those scenarios using automated backup into Amazon Simple Scalable Storage (S3) storage and frequent commits to GitHub. Both services offer industry-standard durability, versioning capabilities, encryption at rest, and authentication controls.

The researcher is responsible for building the artifacts, measuring their accuracy, and reporting the results. There is the potential for biases impacting the study due to resource constraints. For instance, the project might plan four kinetic-700 labels but only three work successfully. In that case, the results should not ignore the failure and instead discuss potential reasons for the issue. It is beyond this project’s scope to validate every situation though it should make reasonable attempts. Additionally, controls are in place to limit cheating or deceiving the results. For example, the result data originates through an automated process.

## Summary

The problem to be addressed in this study is the inability of elderly and special needs care organizations to capitalize on the effectiveness and efficiency of autonomous assistants for human activity recognition. These challenges exist because it’s difficult for researchers to experiment within personal private space. This study aims to mitigate those issues using open-source videos of real people.

The study uses industry-standard tools (e.g., Amazon SageMaker) to build a human activity recognition (HAR) computer vision (CV) model. Those tools include built-in instruments and metrics for assessing the machine-learning process's performance, quality, and efficiency. The study procedure will generate randomized worlds for the characters to move around and utilize virtual cameras to monitor their behaviors. Next, a data analysis phase will validate that the HAR model can reliably extract intents (RQ1) and the process's efficiency (RQ2). Numerous assumptions and delimitations influence this project. First, that modeling HAR is possible within the virtual environment. Another hypothesis is that this approach is even a practical research method. Delimitations are necessary due vast combinatorial set of behaviors. Examining every potential aspect of this problem is impossible, so prioritization is essential. Lastly, the study lacks ethical or privacy concerns because all video recordings are public.

This chapter outlines the research method to build a human activity classification model by examining skeletal movements and object detection overlays using the kinetic-700 dataset. It details the mechanism for metadata extraction and the strategy for converting this data into movement signatures. The next chapter provides the findings from implementing this artifact and evaluating thousands of YouTube videos across the cloud computing environment.

# Chapter 4: Findings

The problem to be addressed in this study was the inability of elderly and special needs care organizations to capitalize on the effectiveness and efficiency of autonomous assistants (Blackhurn, 2021; Kim & Kim, 2021). This constructive research study provides an understanding of the effectiveness and efficiency of AI/ML-based assistants for detecting patient behaviors for improving elderly and special needs care organizations. These situations have a high barrier to entry in studying due to technical constraints, limitations in reproducing results, and privacy and safety concerns. It delivers this capability by modeling human movements within labeled video recordings by tracking the subject’s skeletal movements.

This chapter outlines the experiment’s findings and answers to the research questions. Specifically, this study aims to understand the following two aspects of autonomous assistant effectiveness and efficiency:

### RQ1

What is the effectiveness of autonomous assistance for classifying behaviors of elderly and special needs patients for care organizations?

### RQ2

What is the efficiency of autonomous assistance for classifying behaviors of elderly and special needs patients for care organizations?

## Validity and Reliability of the Data

The experiment ensures data quality by adopting the kinetics 700-2020 data set from DeepMind. “Kinetics is a large-scale, high-quality collection of 650,000 YouTube video clips that cover sixty-five thousand human action classes” (DeepMind, 2020, p.1). Humans manually annotated ten-second segments with single action classes such as playing instruments, shaking hands, and jumping. Alphabet, the parent company of DeepMind and Google, has vetted labeling accuracy through human content moderation and statistical automation (Smaria et al., 2020). The NCU Library search engine results contain at least twenty-two publications that cited this data set and successfully leveraged it for their research.

The kinetics dataset has four core advantages: credibility, transferability, dependability, and confirmability. This study’s results are more credible because they must align with existing survey publications, such as Zhu et al. (2021) and Orhan (2021), or have a strong justification. The dataset also has the potential to gain transferability due to its usage of real-world people in realistic scenarios. Synthetic data must statistically model natural outcomes while preserving characteristics of real-world clinical cohorts (Tian et al., 2018). Each action class has at least seven hundred examples, improving the likelihood of the predictions’ rigor over smaller self-made datasets (Klem et al., 2022). Lastly, the results are more confirmable because external parties choose the videos, and they are publicly accessible. This property also limits the researcher’s ability to cherry-pick or insert biases.

## Results

The kinetics-700 training dataset contains 530,510 YouTube videos that third-party users have uploaded. This analytics pipeline successfully downloaded this data set into an Amazon S3 bucket (9.9TB). The first attempt retrieved 424,613 videos (80%), with most failures due to YouTube service throttling. Since the architecture implements checkpointing scheme, the subsequent retrieval requests skip completed download tasks. This effective strategy helped cache the data set locally and minimized the network I/O requirements. A set of descriptive statistics that map the videos to labels is available in the following table (see Table 12).

**Table 12***Processed Video Category Statistics*

|  |  |
| --- | --- |
| Statistic | Value |
| Total Categories | 700 |
| Minimum Videos Per Category | 250 |
| Maximum Videos Per Category | 881 |
| Standard Deviation Videos Per Category | 128 |
| Median Videos Per Category | 611 |
| 95th Percentile Videos Per Category | 838 |
| 99th Percentile Videos Per Category | 860 |

**Figure 33***High-Level Analysis Process*

The kinetic-700 dataset’s annotations specify the label, time offset, and duration of the target action. A custom video pipeline used the OpenCV library to sample one frame every half-second of each clip (see Chapter 3: Study Procedure section; Figure 33). Labeled segments are at most ten seconds resulting in up to 20 frames/video. Amazon Elastic Container Service (ECS) scheduled the library operations across 38 x Amazon EC2 p4gdn.xlarge instances for 49 hours (152 VCPU, 608GiB RAM, and 38 NVIDIA T4 GPUs). This research project selected Amazon EC2 Spot instances, ephemeral cloud computing resources with up to 90% cost savings.

The Amazon ECS cluster processed 4.2 million seconds of video during the experiment using 6.7 million computations seconds. All code within the cluster emits telemetry to AWS X-Ray, a distributed tracing solution. These traces report that processing high-resolution (1080p) clips takes 9.67 seconds, with 86% of the time spent waiting on network I/O (checkpointing frames). Optimizing the checkpoint logic reduced the average per-video processing time to 1.94 seconds. The entire data set could reprocess in 0.94 million total computation seconds. Suppose the video processor supported multi-threading. In that case, the cluster size could process one video per core and complete the extraction in 1.72 hours.

The OpenPose framework inferred millions of potential human poses within frames as 25x3 matrices. Each item represents a likely body and the location of its twenty-five body parts (see Figure 34). The framework reports body part locations as the three-part tuple X, Y, and confidence score. OpenPose doesn’t provide any consistent list order guarantees of potential humans detected. For instance, the first frame might report Alice, Bob, and Charlie – versus the second reports Bob, Charlie, and Alice. A heuristic algorithm compared the relative movement distance of skeletons across frames to provide consistent sorting (see *Chapter 3: Preprocessing the data*).

Next, a normalization and annotation process assessed the motion sequences. This process began with encoding the motion sequences into Nx25x3 matrics, where *N* represents the subsequent half-second frames. Since videos originate in different resolutions and frame sizes, the body part locations’ X&Y-coordinates became regularized from absolute pixel offsets into relative distances between zero and one. This step aims to enable motion sequence comparability across discrete videos.

**Figure 34***Pose Output Format Body-25*

A picture containing diagram

Description automatically generated

Amazon Rekognition, a computer vision service, further annotated the frames with object, activity, and facial detection metadata. This information came from a post-processing Amazon S3 Batch Job that iterates across the Frame Store and passes metadata to a custom Amazon Lambda function. Initially, the Amazon Rekognition request rate limits caused the batch jobs to terminate unsuccessfully. Geo-distributing the function’s traffic across the AWS cloud mitigated these issues by increasing the service quota 12x (see Figure 35). A copy of the Amazon Rekognition service responses exists in Amazon S3 for future research reproducibility.

**Figure 35***Geo-distributing Traffic*

valid\_regions = [

'us-east-1', 'us-east-2', 'us-west-1','us-west-2',  
 'eu-central-1','eu-west-1','eu-west-2', 'ap-south-1',  
 'ap-northeast-2','ap-southeast-1', 'ap-southeast-2',   
 'ap-northeast-1']

def \_\_detect\_labels\_with\_retry(self, \*\*kwargs)->dict:

for \_ in range(0, 5):

try:

region = valid\_regions[randint(0,len(valid\_regions)-1)]

rekognition = boto3.client('rekognition', region\_name=region)

print('DetectLabels(%s) - s3://%s -> %s' % (

region,

self.manifest.report.frame\_bucket,

self.manifest.video\_id

))

return rekognition.detect\_labels(\*\*kwargs)

except ProvisionedThroughputExceededException as error:

print('ProvisionedThroughputExceededException -- %s' %

str(error))

sleep(randint(10,50)/10)

raise Exception('Unable to detect\_labels - %s' %   
 self.manifest.video\_id)

Downloading and extracting metadata from the videos produced three manifest reports per video (1,591,530 files). This corpus represents the facts and evidence to address this dissertation’s research questions. Specifically, what are autonomous agents' effectiveness (RQ1) and efficiency (RQ2) in assisting elderly and special needs care facilities? Before transforming these facts into answers, this research project needed to overcome a big data problem. Each manifest file references annotated frames, skeletal positioning, prediction confidence vectors, and object detection labels. These 21.7 million semi-structured documents span Amazon S3 buckets, DynamoDB tables, and Elastic FileSystem network storage.

GraphQL is a declarative data-fetching method that enables web clients to describe the capabilities and requirements of data models (GraphQL, 2021). A server endpoint fulfills the request using *resolvers* that retrieve entity definitions from arbitrary data stores. For instance, a client requests the total number of people in each video frame, the annotation metadata, visible body positions, and people identifiers (see Figure 36). Internally, the GraphQL service determines the response requires the resolver for *annotation* and *analysis* base entities. After binding these entities, the child resolvers execute to fetch *frames* metadata and recursively acquire the frame’s *bodies* results.

This constructive research project built the GraphQL service using Amazon AppSync and Amazon Lambda functions. AppSync is responsible for processing the queries and orchestrating the business logic to fetch information from the various data stores (e.g., DynamoDB and S3). This capability surfaced consistency issues, accelerated development, and streamlined data retrieval because of the uniform access to information without exposing internal serialization, partitioning, and database technology. Without these features, significant investments are necessary to clean, catalog, and consolidate the data ahead of time. That would introduce risk to the project and its finite timeline

**Figure 36**  
*Example GraphQL Query*

{

get\_video(video\_id: “---0dWlqevI”) {

annotation {

label

}

analysis {

frames {

offset

people\_count

bodies {

rshoulder {

visible

}

identity {

person\_id

}

}

}

}

}

}

An excerpt of the previous query is available in the following figure. It’s worth noting how the response aligns with the data model request and gives the consumer a single combined document. In contrast, similar technologies like REST (REpresentational State Transfer) would require the caller to parse multiple responses and combine them.

**Figure 37**  
*GraphQL Response*

{

"data": {

"get\_video": {

"annotation": {

"label": "clay pottery making"

},

"analysis": {

"frames": [

{

"offset": 19,

"people\_count": 1,

"bodies": [

{

"rshoulder": {

"visible": true

},

"identity": {

"person\_id": 0

}

]

}

]

}

}

}

### RQ1

*What is the effectiveness of autonomous assistants for classifying behaviors of elderly and special needs patients for care organizations?*

Effectiveness is the degree to which something successfully produces a desired (Oxford, 2023). This constructive research project reliably extracts metadata from video sequences and surfaces that information into an extensive schema. The initial implementation scope focuses on human identification, tracking, and annotating capabilities. Future research efforts can quickly extend the feature set to add domain-specific classification labels. For instance, an elderly care facility could include a fall detection algorithm powered by these foundational properties.

The implementation utilizes a loosely coupled analytics pipeline that first identifies the humans and their skeletal positions within frames using the OpenPose framework. Next, a custom Movement Tracker reliably determines the motion sequence for each person across the sampled clip. Third, Amazon Rekognition further annotates those frames and each person’s bounding boxes with object detection. The amalgamation of these capabilities provides greater predictive power than any single component. For example, the OpenPose framework offers a foundational ability to extract skeletal positions from a 2-D frame. This information is sufficient for differentiating core movements such as walking, throwing, sitting, and eating. However, predicting many derived activities from only skeletal movements is challenging. For instance, *playing cello* and *playing clarinet* have similar action sequences due to sharing a parent activity (*playing an instrument*). Similarly, object detection can predict that a cello is within the player’s bounding box but cannot decern if it’s in use. This analytic pipeline successfully composites that the person is *playing an instrument* and *the instrument is a cello* (see Figure 38).

**Figure 38***Playing the cello (Video: -23ykna85DI)*

A group of people playing instruments on a stage

Description automatically generated with medium confidence

The previous figure illustrates the results from OpenPose and contains the predicted skeleton of the cello player. Similarly, the predictive labels from Amazon Rekognition are available in the proceeding figure with clues that this is an Adult, Performer, Musical instrument, Cello, Person, and Solo Performance (see Figure 39). Policy engineers can quickly and consistently utilize this information for automated reasoning systems. Suppose a special needs facility has children and adult patients. In that case, policies could exist to flag children drinking beer as requiring remediation but permit the adults. Systems engineers could codify the policy requirements into the GraphQL analysis schema as the *underage\_drinking* flag. Recursively, the drinking flag can integrate into more sophisticated policies. Ultimately, this means that the system can inherently predict many characteristics of video clips, and it’s a straightforward process to extend the schema to incorporate additional domain-specific detections.

**Figure 39***Cello with label annotations*

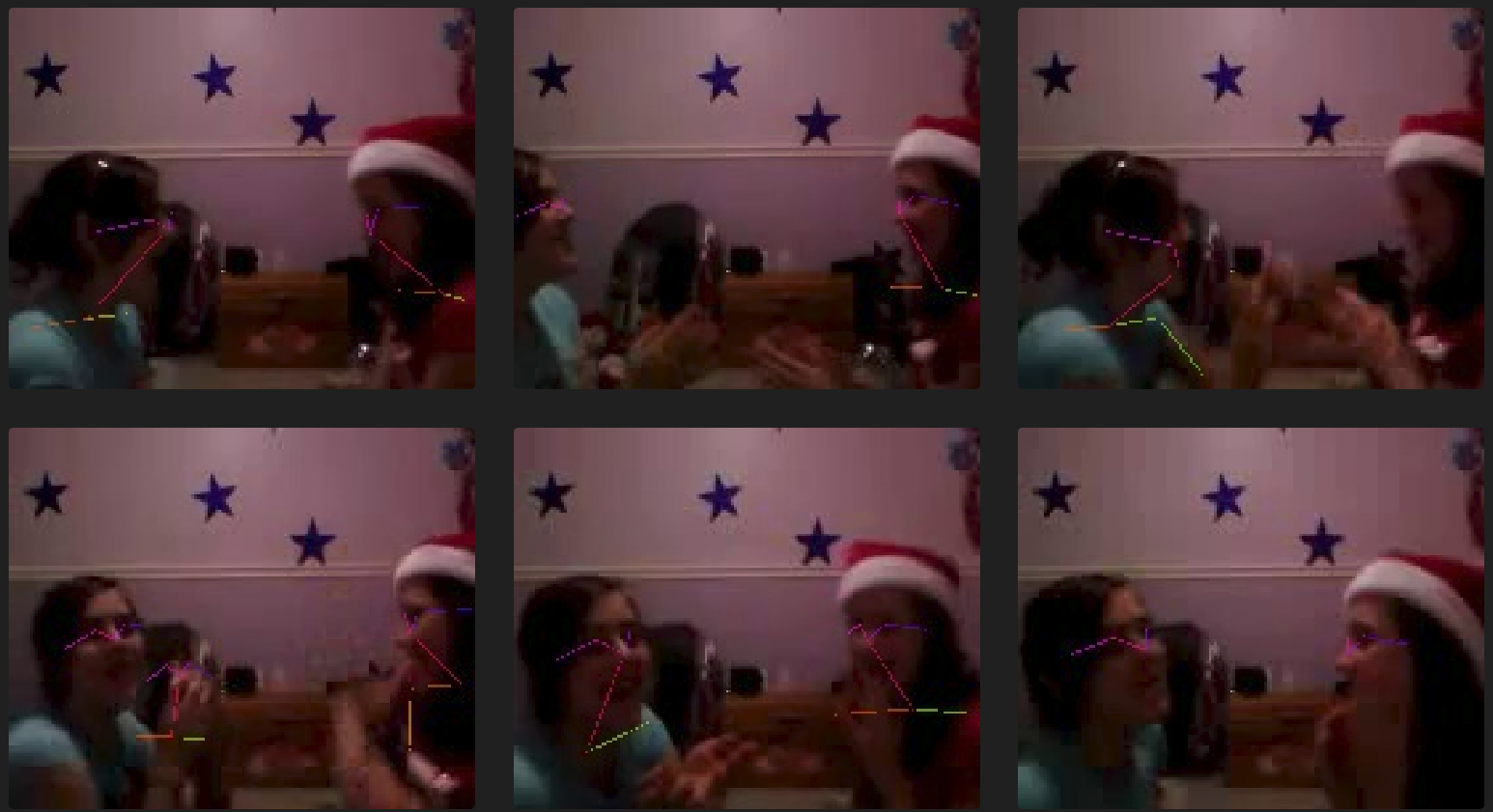
A screenshot of a computer

Description automatically generated with medium confidence

Using full-body skeletal monitoring is also insufficient for several kinetic-700 action types. For instance, numerous videos within the categories of *waving* and *washing hands* focus the camera on only the person’s hands. This situation causes OpenPose not to detect other body parts and return a low-confidence 25x3 position matrix. Carnegie Mellon’s team has addressed this situation with two purpose-built models for faces and hands (Hidalgo et al., 2019). The analytics pipeline could introduce a classification step to visible body type (e.g., whole body, hands-only, face-only) based on preliminary investigations. This detection could inform the system how to parse skeletal metadata correctly. However, a detailed analysis of this property is outside the research’s scope.

Another set of challenges arises with low-resolution mobile phone recordings and other blurry motion captures. These situations cause the OpenPose framework to predict phantom limbs and bogus skeletal matrics, impacting automated analysis (see Figure 40).

**Figure 40***Playing hand-clapping games (Video: -MOpSXQ5ZcU)*



Requiring more information than simple skeletal movements was expected and called out during the literature review (see *Chapter 2: How does human activity recognition work*). The research project increased accuracy by adding object detection metadata to the frames using Amazon Rekognition. There are several benefits to using this service, such as it’s a RESTful endpoint that doesn’t require administrative overhead. For example, their body posture would infer that the person is sitting in a chair and labels for the *instrument* and *cello* detections within the frame. Calculating the likelihood of the person using the object was straightforward, using collision detection with the human’s and object’s bounding boxes. While this methodology was broadly effective, it didn’t work with perspective shots. For instance, if the cello were near the camera, its bounding box would saturate the frame and trigger the collision logic.

### RQ2

*What is the efficiency of autonomous assistance for classifying behaviors of elderly and special needs patients for care organizations?*

Efficiency is the quality of doing something well without wasting time or resources (Oxford, 2023). An underlying assumption of this research project is autonomous assistance can monitor the person in their home and replace nursing staff with electronics. These video recording devices have unfiltered access to the patient’s privacy and data security. This situation makes it ideal for the analysis to occur on-premises and never leave the local network. Additionally, this design improves the system’s reliability by removing remote dependencies. However, it also means that the system must have high efficiency to perform these operations using commodity hardware.

The Amazon ECS cluster processed 4.2 million seconds of video during the experiment using 6.7 million computations seconds. All code within the cluster emits telemetry to AWS X-Ray, a distributed tracing solution. These traces report that processing a high-resolution (1080p) clip takes 9.67 seconds, with 86% of the time spent waiting on network I/O (checkpointing frames). Suppose a reduction or elimination of these checkpoints occurred. In that case, the entire data set could reprocess in 0.94 million computation seconds. There are further potential performance gains from multi-threaded extraction processes. This outcome means the approach is sufficiently efficient before any optimizations.

The complete dataset ballooned to 21.7 million documents across 9.9TB of storage through an unintentional side-effect of this research project. However, the GraphQL interface can efficiently fetch and map frame-level results into arbitrary data models within single-digit seconds. This construct utilizes a fully serverless design pattern that can scale elastically to any traffic size and provide consistent performance. These characteristics are ideal for expanding the project into multi-tenant use cases.

Similarly, the loosely coupled analytics pipeline components implement autoscaling policies based on usage patterns. For instance, the OpenPose framework performs best on compute instances with GPU acceleration cards. The system minimizes costs by queuing requests to those instances until Amazon EC2 Spot instances become available (up to 90% savings). When the frame analysis is complete, the report is JSON-encoded and written to Amazon S3, triggering downstream processors to reach an eventual consistency state. Suppose a caller requests details that aren’t yet available. In that case, the GraphQL interface returns a partial response and flags denoting to reattempt later.

## Evaluation of the Findings

Design science research is one of the most common research methods for information systems and technology (Silvestrini & Sammito, 2012). These studies identify a problem, build artifacts, and communicate the implementation’s unique value (Hevner et al., 2004). In addition, many researchers follow this process to build proof-of-concept and execute case studies.

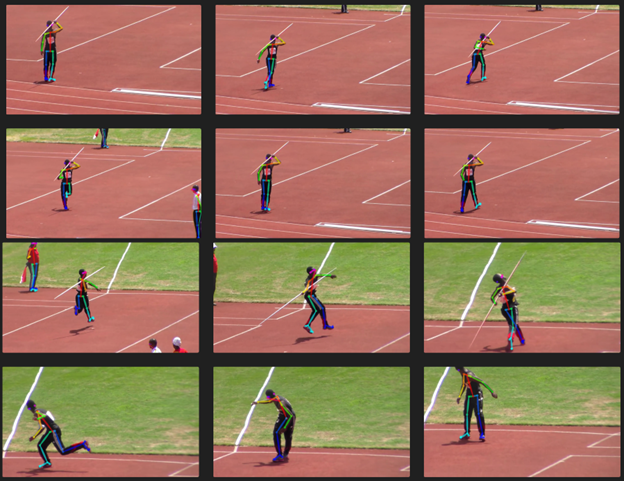
Within this study’s context, this meant creating an analytics pipeline that extracts metadata from videos and presents that information to policy engines that help elderly and special needs patients. The research project demonstrates that it’s possible to integrate loosely coupled frame and motion analyzers into a unified extensible schema. This capability enables practitioners to build domain-specific detection logic that reuses foundational features such as identity, motion tracking, and object detection. For instance, the detection taxonomy could contain medication labels combing *eating*-sequence with *medication-*detected flags. Furthermore, entrepreneurs can package these solutions onto commoditized hardware and know sufficient processing power exists to keep data local and secure. While this version utilizes Amazon Rekognition, creating offline object detection models using OpenCV or another industry-standard technology would be possible.

These findings align with recent state-of-the-art publications and classical theory. Ballard and Zhang (2021) stated that primate vision relies on a hierarchical system of understanding and that system layers annotations to derive deeper insights. Likewise, this research found many indoor activities have a high similarity score. Das et al. (2016) enumerated challenges to HAR predictions within indoor environments, such as many activities having similar signatures. For example, a person in a sitting position could be eating, watching tv, or reading a book. These situations share a partner action of *seating,* and the derived behavior becomes apparent only through annotations.

Meanwhile, outdoor activities are more expressive and pronounced signatures. The first six frames are the composite activity of *holding something* and *standing*. During the subsequent six frames, the entire body rotation is an entirely different activity – *throwing* (see Figure 41). This activity logically makes sense as people set up, perform, and conclude action sequences. It also illustrates how several actions are classifiable through composite action classification models. Amerineni et al. (2021) recommended a similar mechanism that utilizes seven classification models to score 18 punches and 24 kicks. While this approach is intuitive, it’s not the default solution. For instance, Anderson et al. (2022) and Friedman et al. (2023) recently chose to constrain the action space and focus on purpose-built solutions. Both research groups identify design simplicity, data limitations, and the commercial benefits of purpose-built algorithms.

Within the kinetic-700 data set, actors transitioning between distinct actions are surprisingly common. The loosely coupled analytics pipeline and GraphQL interface could handle these scenarios by integrating simple detectors. As explained in section *Results RQ1*, the policy engine recursively leverages its knowledge to derive greater specificity in the prediction.

**Figure 41**  
*Javelin Throwing Setup (Video: zVlBFLFkUNk)*



## Summary

This research project extracted metadata from hundreds of thousands of YouTube clips and used that information to evaluate seven hundred discrete actions. Based on those findings, the evidence suggests an efficient and effective mechanism exists for classifying human activity recognition within video sequences. Medical care facilities could leverage these capabilities to monitor their patients securely while maintaining privacy requirements.

Realizing this outcome is possible through an extensible schema that goes beyond root behaviors like seating, standing, and running. Object detection and image annotation can provide sufficient information to derive child activities like playing soccer versus basketball. The domain-specific detection logic expands to an arbitrary depth through additional levels of annotations. After collecting enough information, commoditized hardware can economically run computer vision models within patients’ homes. This approach restricts the data movement within the local network and removes the need to exfiltrate sensitive information for analysis.

However, there are still challenges and limitations to this data structure. For example, many action sequences consist of discrete segments and transitions. Approaching the problem as micro-actions could permit encoding the action space as tiny self-contained behaviors. That would open the door to association rule mining like apriori algorithms. The next chapter presents the future implications, recommendations, and conclusions.

# Chapter 5: Implications, Recommendations, and Conclusions

The problem to be addressed in this study was the inability of elderly and special needs care organizations to capitalize on the effectiveness and efficiency of autonomous assistants (Blackhurn, 2021; Kim & Kim, 2021). This constructive research provides an understanding of the effectiveness and efficiency of AI/ML-based assistants for detecting patient behaviors for improving elderly and special needs care organizations. These situations have a high barrier to entry in studying due to technical constraints, limitations in reproducing results, and privacy and safety concerns. It delivers this capability through an extensible schema and CV-based features for identity, movement tracking, and object detection.

Design science is a research methodology that creates and uses purposeful artifacts to study a phenomenon (Hevner et al., 2004). Academic and business communities employ this method as a standard approach to information technology and communication (IT&C) problems (Bryar & Carr, 2021; Peffers et al., 2007). The methodology comes with well-defined guidelines to implement a three-phased procedure. First, the researcher(s) must identify a domain-specific challenge. Next, that researcher creates artifacts that study this phenomenon. Third, those artifacts assess the topic and communicate answers to the research questions.

This dissertation examines human activity recognition (HAR) for elderly and special needs patients. It leverages a loosely coupled analytics pipeline to process the DeepMind’s kinetic-700 data set, containing 650,000 YouTube videos with ten-second labeled segments. The pipeline extracts metadata into multiple NoSQL databases using the OpenPose Framework, Amazon Rekognition, and custom Python logic. Finally, users can quickly retrieve those results using a GraphQL interface. These artifacts provide data consumers with a near real-time query interface that’s straightforward to extend with domain-specific requirements.

The final chapter contains an examination of the implications of Chapter 4’s findings and their potential influence on future commercial and academic investments. It also frames this project’s learnings in the broader context of the doctoral body of knowledge and the literature review.

## Implications

Three converging macro-tends are an increasing elderly population, an increasing cost of assisted living, and a decreasing nursing population. This combination means that fewer people will have access to quality care facilities in the future. Patients require mechanisms that allow them to remain in residence longer and acquire benefits from nursing staff through automation. One solution is to deploy low-cost cameras that monitor the patients and make recommendations based on their behavior. However, building and operationalizing those capabilities is challenging because of personal privacy and security concerns. This research project mitigates these issues by utilizing public videos.

### RQ1

*What is the effectiveness of autonomous assistants for classifying behaviors of elderly and special needs patients for care organizations?*

Effectiveness is the degree to which something successfully produces a desired. Within this research context, that means that the analytics pipeline can reliably download, extract, and index kinetic-700 videos (see *Chapter 4: Findings, RQ1*). The pipeline invoked the loosely coupled components resulting in 21.7M prediction documents describing the state within sampled frames (2 fps). These predictions identify distinct people within the video, movement tracking, and object detection features. The GraphQL interface decouples the data retrieval from the data storage constructs.

This configuration enables future data scientists to add CV models dynamically and bring new foundational capabilities. For example, a data scientist could use TensorFlow or PyTorch to build a medication object classification model. It also permits developers to extend the GraphQL schema by rule mining and through business heuristics requirements. For instance, they could detect the presence of medicine and the subject eating something. The amalgamation of these insights and detections permits the system to be reasonably confident the patient is taking their medication.

These outcomes are broadly expected and align with the literature review. For instance, Zhang et al. (2021) used a similar technique for clustering movements unsupervised. Das et al. (2018) also demonstrated HAR predictive capabilities using recursive neural networks (RNN) algorithms. Where this research project’s implementation diverges is the mechanism predicting the action. Instead of creating one sophisticated classification algorithm, it relies on an ensemble of dumb signals. This strategy resembles Bell et al.’s (2009) Netflix Prize solution that combines 107 trivial predictors into one high-precise recommendation engine. One of the core strengths of ensemble algorithms is the ability to mature the various signal predictor independently. That design pattern is beneficial for addressing some of the challenges and limitations of this project’s version. For example, several videos zoom in on the actor’s hands, confusing the model for detecting 25 body points. Preprocessing actions could exist to classify the shot type and then decide between the body, hands, and face models. These minor insertions into the analytics pipeline also enable the predictions to support specific markets.

### RQ2

*What is the efficiency of autonomous assistants for classifying behaviors of elderly and special needs patients for care organizations?*

Efficiency is the quality of doing something well without wasting time or resources. Within this research context, it means processing frames quickly and supporting an economically elastic computation scale-out model. The analytics pipeline can download, extract, and index metadata in near real-time utilizing a GPU farm. This farm leverages a task-queuing strategy that waits for the availability of inexpensive computation resources. That characteristic was essential for this low-budget project. However, enterprise environments could adjust their pricing tolerance to meet customer expectations. After allocating those resources, the pipeline churns through video clips efficiently. Additional performance improvements are possible due to the Write-Once Read-Maybe (WORM) semantics. Many WORM computations are deferrable until the GraphQL interface requests the data. This optimization further reduces the cloud resources necessary to process the videos by deferring computations that aren’t accessed.

**Figure 42***Pouring beer (-5HJWCQ02Ds)*

A picture containing text

Description automatically generated

It’s challenging to reliably extract details from every video because many activities look similar. For example, consider the hand movements necessary to pour a beer versus milk (see Figure 42). These two actions derive from a common ancestor, and what makes them distinct is the specific object liquid. This situation is a perfect example of the benefit of overlaying multiple CV models. For instance, the skeletal movement model can recognize the *pouring-liquid* action. Secondary computer vision (CV) systems can perform object detection to predict beer versus milk. A tertiary CV source could utilize a thermal camera to support the derived action of *pouring hot milk*.

Furthermore, it’s possible to introduce derived activities based on the labels from these model predictions. Suppose one wants to model when the patient goes to sleep. In that case, the system can factor into its calculus the patient is drinking warm milk or consuming cold beer, impacting their sleeping schedule.

## Recommendations for Practice

This dissertation examines human activity recognition using Deep Mind’s kinetic-700 dataset as a proxy for elderly and special needs behaviors. However, the capability broadly applies to markets that demand quality control and oversight. For instance, parents would likely pay a premium to daycares and school systems that can summarize their child’s positive and negative interactions. The court systems use naïve ankle monitoring because continuous video monitoring would violate personal privacy. However, under the proposed implementation, tracking the defendant's actions and producing a secure log of behaviors is possible (Mickens, 2018). Also, numerous situations exist where public safety has priority over personal privacy, such as sporting events, public events, transportation, and prisons. There are commercial applications outside of continuous personal monitoring scenarios. For instance, manufacturing and assembly lines can monitor for health and safety risks.

## Recommendations for Future Research

Future researchers can verify the results of this dissertation and quickly extend them within their private AWS accounts. The solution is fully automated using the AWS Cloud Development Kit (AWS CDK v2), which provisions an identical environment with minimal effort. The source code is available at <https://github.com/dr-natetorious/Dissertation>. Follow the README.md documentation for deployment instructions and troubleshooting guidance.

This constructive research project examined a broad set of videos and extracted details about a diverse population. However, there is significant potential for improvements across the value chain. For example, the video analysis uses a fixed frame rate (2 fps) for all clips. No attempt to quantify a reasonable frame rate occurred during this project. Future research should determine the appropriate sampling necessary to make informed decisions.

The OpenPose framework also returns low-quality predictions when overlapping people exist in a small shot (see Figure 43). In these sequences like this clip, the camera pans and rotates as the children dance in a circle. Commercial libraries exist for path tracking within the video and could provide better support than the default OpenPose constructs. A cursory exploration shows these features support limited movements and are an open research topic. One potential solution is to leverage multiple cameras and calculate the 3-D space. That approach would allow depth-specific bounding boxes and improve collision detection with the scene’s objects.

**Figure 43***Clapping (0G1OirEz2OA)*



## Conclusions

Growing macroeconomic trends necessitate transitioning skilled human tasks to autonomous agents. These trends are evident in medical care facilities where the costs are increasing while the availability of trained professionals is decreasing. Meanwhile, automation and computer-aided capabilities are becoming ubiquitous. This situation requires research into autonomous agents that can facilitate tasks traditionally reserved for humans, like nursing. One of the most natural agent interaction models is through computer vision because body language is broadly universal. For example, when a person has a fall, their movement follows a similar signature regardless of nationality. The same statement is true regarding other actions like seating, running, jumping, and standing. However, building and training these computer vision-based agents has a litany of technical and practical challenges. First, researchers must provision enough cameras to collect the required training data, introducing scalability and economic constraints. Next, the patients are unlikely to permit researchers to record them 24/7. Third, other researchers still have issues reproducing snowflake data sets, assuming sufficient patients grant permission and forgo personal privacy. These limitations establish the need for training computer vision models with open-source video repositories.

This constructive research project built an AWS cloud analytics pipeline to process the kinetic-700 dataset. It reliably extracts foundational insights per frame and enables querying them through an industry standard GraphQL interface. These artifacts allow researchers to build sophisticated products for improving elderly and special needs care. It permits data scientists to extend the analytics pipeline with additional CV models. Similarly, developers can extend the GraphQL schema by defining domain-specific aggregate and composite states. This approach aligns with the literature review and goes beyond existing strategies. For instance, most practitioners create purpose-built models constrained to a finite action space. Instead, this research supports arbitrary extensions into the retrieval schema.

This study defined success criteria regarding the effectiveness (R1) and efficiency (R2) of identifying human behaviors within open-source videos. These two related questions aim to measure the accuracy and scalability of the proposed solution. Extracting skeletal movements from video demonstrated an ability to identify several foundational action sequences correctly. After enriching those frames with object detection metadata, the predictive system could handle an arbitrary depth of derived actions, each layer describing greater specificity. This metadata ensembling results in a highly efficient scalable design that future investments could mature to support the general corpus of human behavior. Furthermore, this research project demonstrated that these predictive capabilities could run on commoditized local hardware. That flexibility reduces personal privacy concerns since video recordings never leave the patient’s control.

The literature review paints a picture of researchers progressing into highly sophisticated deep-learning models. These technologies are compelling and will shape the future of AI, especially for generative content scenarios. However, this research demonstrates that simple designs with several dumb signals can still produce high-quality results. Each of these signals can use artificial intelligence. However, it’s critical not to forget that the system must solve a business problem. Within a business problem context, known rules can be codified directly in traditional programming languages. System designers must be mindful in choosing the right tool for the job, whether simple-branch or ANN/DNN constructs.

# References

Aguida, M., Ouchani, S., & Benmalek, M. (2020). A review on cyber-physical systems. *International Conference on Enabling Technologies* (pp. 275-278). Basque Coast, Bayonne; France: IEEE. doi:https://doi.org/10.1109/WETICE49692.2020.00060

Aihara, S., Kitamura, S., Dogan, M., Sakata, S., Kondo, K., & Otaka, Y. (2021). Patients’ thoughts on their falls in a rehabilitation hospital: a qualitative study of patients with stroke. *BMC Geriatrics, 21*(1), 1-14. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edssjs&AN=edssjs.7A908AA7&site=eds-live&scope=site

Amerineni, R., Gupta, L., Steadman, N., Annauth, K., Burr, C., Wilson, S., . . . Vaidyanathan, R. (2021). Fusion models for generalized classification of multi-axial human movement: Validation in sport performance. *Sensors, 24*, 1-10. doi:10.3390/s21248409

Anderson, W., Choffin, Z., Jeong, N., Callibhan, M., Joeng, S., & Sazonav, E. (2022). Empirical study on human movement classification using insole footwear sensor system and machine learning. *Sensors, 22*(7), 2743-2763. doi:10.3390/s22072743

Ariely, D. (2009). *Predictably irrational: the hidden forces that shape our decisions.* HarperCollins Publishers.

Asimov, I. (1942). *Runaround.*

Atienza, R. (2018). *Advanced deep learning with Tensorflow 2 and Keras.* Birminghan, UK: Packt Publishing.

AWS. (2021). *AWS RoboMaker*. Retrieved from Amazon Web Services: https://aws.amazon.com/robomaker/

AWS. (2023). *What is Amazon Simple Queue Service*. Retrieved from AWS Developer Guide: https://docs.aws.amazon.com/AWSSimpleQueueService/latest/SQSDeveloperGuide/welcome.html

AWS. (n.d.). *What is AWS Fargate*. Retrieved from AWS Developer Documentation: https://docs.aws.amazon.com/AmazonECS/latest/userguide/what-is-fargate.html

Ballard, D., & Zhang, R. (2021). The hierarchial evolution in human vision modeling. *Topics in Cognitive Science, 13*(2), 309-328. doi:https://doi.org/10.1111/tops.12527

Barua, H., & Mondal, K. (2019). A comprehensive survey on cloud data mining (CDM) frameworks and algorithms. *CM Computing Surveys. Sep2019, Vol. 52 Issue 5, p1-62. 62p*, 1-62.

Bell, Koren, & Volinsky. (2009). *The BellKor solution to the Netflix Prize.* Retrieved from Netflix Prize: https://netflixprize.com/assets/GrandPrize2009\_BPC\_BellKor.pdf

Besada, D. E. (2020). Resource requirements for community-based care in rural, deep-rural and peri-urban communities in South Africa. *PLoS ONE, 15*(1), 1-19. doi:https://doi.org/10.1371/journal.pone.0218682

Bipin, K. (2018). *Robot Operating System Cookbook.* Packet Publishing.

Blackhurn, B. (2021). Sensitive situations in a nurse residency program: balancing confidentiality with meaningful solutions. *Journal for Nurses in Professional Development, 37*(3), 185-187. doi:10.1097/NND.0000000000000694

Boorugu, R., & Ramesh, G. (2020). A survey on NLP based text summarization for summarizing product reviews. *International Conference on Inventive Research in Computing Applications* (pp. 352-356). Coimbatore, India: IEEE. doi:https://doi-org.proxy1.ncu.edu/10.1109/ICIRCA48905.2020.9183355

Brown, T. (2015). A Primer on data security. *CPA Journal May Volume 85, Issue 5*, 58-62.

Bryar, C., & Carr, B. (2021). *Working backwards: insights, stories, and secrets from inside Amazon.*

Buchanan, B. (2005). A very brief history of artifical intelligence. *AI Magazine, 26*(4), 53-60. Retrieved from ttps://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=ofs&AN=501189619&site=eds-live

Burr, V. (2015). *Social constructionism.* Routledge.

Cao, Z., Hidalgo, G., Simon, T., S, W., & Sheikh, Y. (2021). OpenPose: realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172-186. doi:10.1109/TPAMI.2019.2929257

CDC. (2016, July 6). *Motor vehicle crash deaths*. Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/vitalsigns/motor-vehicle-safety/index.html

Centers for Disease Control and Prevention. (2020). *Centers for disease control and prevention, National Center for Injury Prevention and Control. Web-Based Injury Statistics Query and*. Retrieved from Centers for Disease Control and Prevention: http://www.cdc.gov/injury/wisqars

Chen, B., & Curtmola, R. (2017). Remote data integrity checking with server-side repair. *Journal of Computer Security 25*, 537-584.

Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2018, January). Model compression and acceleration for deep neural networks. *IEEE Signal Processing Magazine, 35*(1), 126-136. doi:https://doi-org.proxy1.ncu.edu/10.1109/MSP.2017.2765695

CMU. (2021). *CMU graphics lab motion capture database*. Retrieved from Carnegie Mellon University: http://mocap.cs.cmu.edu/

Cohen, B. (2013, November 14). *The 10 smartest cities In North America*. Retrieved from Fast Company: https://www.fastcompany.com/3021592/the-10-smartest-cities-in-north-america

Commission of the European Communities. (2009). *Internet of Things — an action plan for Europe.* Retrieved from http://eurlex.europa.eu/LexUriServ/site/en/com/2009/com2009\_0278en01.pdf

Darwin, C. (1859). *On the origin of species.*

de Waal, D., & du Toit, J. (2011). Automation of generalized additive neural networks for predictive data mining. *Applied Artificial Intelligence, 25*(5), 380-425. doi:10.1080/08839514.2011.570156

DeepMind. (2020). *Kinetics*. Retrieved February 11, 2023, from DeepMind: https://www.deepmind.com/open-source/kinetics

Denis, D. (2015). *Applied univariate, bivariate, and multivariate Statistics* (1st ed.). John Wiley & Sons, Incorporated.

Dickson, A., Emad, H., & Adu-Agyum, J. (2018). Theoetical and conceptual framework: mandatory ingredients of quality research. *International Journal of Scientific Research, 7*, 438-441. Retrieved from https://www.researchgate.net/publication/322204158\_THEORETICAL\_AND\_CONCEPTUAL\_FRAMEWORK\_MANDATORY\_INGREDIENTS\_OF\_A\_QUALITY\_RESEARCH

Donovan, C. (2016, August 30). *Power and effect size*. Retrieved from YouTube: https://www.youtube.com/watch?v=9LVD9oLg1A0

Edureka. (2018, October 16). *Natural language processing in 10 minutes*. Retrieved from YouTube: https://www.youtube.com/watch?v=5ctbvkAMQO4

Fedus, W., Zoph, B., & Shazeer, N. (2022). Switch transformers: scaling to trillion parameter models. (A. Clark, Ed.) *Journal of Machine Learning, 23*, 1-40. doi:https://jmlr.org/papers/v23/21-0998.html

Ford, D. (2012, March 18). *As cars are kept longer, 200,000 Is the new 100,000*. Retrieved from The New York Times: https://www.nytimes.com/2012/03/18/automobiles/as-cars-are-kept-longer-200000-is-new-100000.html

Fridman, L. (2017, January 16). *MIT 6.S094: Introduction to deep learning and self-driving cars*. Retrieved from YouTube: https://www.youtube.com/watch?v=1L0TKZQcUtA&feature=youtu.be

Fridman, L. (2020, January). *Deep Learning state of the art*. (Massachusetts Institute of Technology (MIT)) Retrieved from YouTube: https://youtu.be/0VH1Lim8gL8

Friedman, L., Prokopenko, V., Katrychuk, D., & Komogortsev, O. (2023). Factors affecting inter-rater agreement in human classification of eye movements: a comparison of three datasets. *Behavior Research Methods, 55*(1), 417-427. doi:https://doi.org/10.3758/s13428-021-01782-4

Frolov, S. (2021). Quantum computing’s reproducibility crisis: Majorana fermions. *Nature: International Weekly Journal of Science, 592*(7854), 350-352. doi:https://doi.org/10.1038/d41586-021-00954-8

Fu, Z. (2019). An introduction of deep learning based word representation applied to natural language processing. *International Conference on Machine Learning, Big Data and Business Intelligence*, (pp. 92-104). doi:https://doi-org.proxy1.ncu.edu/10.1109/MLBDBI48998.2019.00025

Gan, Q., Li, Y., Wang, G., & Zhang, Y. (2020). Application research of optical tracking point layout in computer motion capture technology. *International Conference on Innovation Design and Digital Technology* (pp. 548-552). Zhenjing, China: IEEE. doi:10.1109/ICIDDT52279.2020.00109

García-Pérez, M. A. (2012). Statistical conclusion validity. *Frontiers in Psychology, 3*. doi:https://doi.org/10.3389/fpsyg.2012.00325

Gergen, K. (2010). *Social constructionist ideas: theory and practice*. (The Taos Institute) Retrieved from Vimeo: https://vimeo.com/15676699

Gorgulu, Y., & Tasdelen, K. (2020). Huamn activity recongition and temporal action localization based on depth sensor skeletal data. *Innovations in Intelligent Systems and Applications Conference* (pp. 1-5). Istanbul, Turkey: IEEE. doi:https://doi-org.proxy1.ncu.edu/10.1109/ASYU50717.2020.9259886

GraphQL. (2021, October). *GraphQL specification*. Retrieved from GraphQL: https://spec.graphql.org/October2021/

Guinness World Records. (2022). *Heaviest man ever*. Retrieved from https://www.guinnessworldrecords.com/world-records/heaviest-man

Hole, H., & Ahmad, S. (2019). Biologically driven artificial intelligence. *Computer, 52*(8), 72-75. doi:10.1109/MC.2019.2917455

Hornberg, A. (2017). *Handbook of machine and computer vision.* John Wiley & Sons, Incorporated.

Huang, M., Rust, R., & Maksimovic, V. (2019). The feeling economy: managing in the next generation of artificial intelligence. *California Management Review, 61*(4), 43-65. doi:https://doi-org.proxy1.ncu.edu/10.1177/0008125619863436

Jackson, B., & Rege, M. (2019). Machine learning for classification of economic recessions. *IEEE 20th International Conference on Information Reuse and Integration for Data Science* (pp. 31-38). Los Angeles, CA, USA: Institute of Electrical and Electronics Engineers. doi:10.1109/IRI.2019.00019

Jaisswal, A., & Naik, A. (2021). Effect of Hyperparameters on Backpropagation. *Pune Section International Conference* (pp. 1-5). IEEE. doi:10.1109/PuneCon52575.2021.9686489

Jason, L., & Glenwick, D. (2016). *Handbook of methodological approaches to community-based research : qualitative, quantitative, and mixed methods .* Oxford University Press.

Kahn Academy. (2014). *Origin of Markov Chain*. Retrieved from Kahn Academy: https://www.khanacademy.org/computing/computer-science/informationtheory/moderninfotheory/v/markov\_chains

Kane, T. (2019, March). Artificial intelligence in politics: establishing ethics. *Technology and Society Magazine, 38*(1), 72-80. doi:10.1109/MTS.2019.2894474

Keller, J., Liu, D., & Fogel, D. (2016). *Fundamentals of computational intelligence.* John Wiley & Sons.

Kilgallon, S., De la Rosa, L., & Cavazos, J. (2017). Improving the effectiveness and efficiency of dynamic malware analysis with machine learning. *Resilience Week* (pp. 30-36). Wilmington, Delaware: IEEE. doi:10.1109/RWEEK.2017.8088644

Kim, J., & Kim, S. (2021). The determinants of caregiver use and its costs for elderly inpatients in Korea. *BMC Health Services Research, 21*(631), 1-10. doi:https://doi.org/10.1186/s12913-021-06677-w

Kim, K., & Cho, S. (2008). Evolutionary ensemble of diverse artificial neural networks using speciation. *Neurocomputing, 71*(7-9), 1604-1618. doi:https://doi.org/10.1016/j.neucom.2007.04.008

Klem, N., Bunzli, S., Smith, A., & Shields, N. (2022). Demystifying qualitative research for musculoskeletal practitioners Part 5: rigor in qualitative research. *Journal of Orthopaedic & Sports Physical Therapy, 52*(2), 60-62.

Krizhevshy, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *25th International Conference on Neural Information Processing Systems* (pp. 1097-1105). Red Hook, NY: ACM. doi:https://dl.acm.org/doi/10.5555/2999134.2999257

Langer, M., He, H. Z., Rahayu, W., & Xue, Y. (2020). Distributed training of deep learning models: a taxonomic perspective. *Transactions on Parallel and Distributed Systems, 31*(12), 2802-2818. doi:https://doi.org/10.1109/TPDS.2020.3003307

Langston, A. (2022, March 14). *In a historic milestone, Azure Quantum demonstrates formerly elusive physics needed to build scalable topological qubits*. Retrieved from Microsoft Innovation Stories: https://news.microsoft.com/innovation-stories/azure-quantum-majorana-topological-qubit/

Lee, J., & Yoo, H. (2021). An overview of energy-efficient hardware accelerators for on-device deep-neural-network training. *Open Journal of the Solid-State Circuits Society, 1*, 115-128. doi:https://doi.org/10.1109/OJSSCS.2021.3119554

Lui, H., K, S., Fernando, C., & Kavukcuoglu, K. (2018). Hierarchical representations for efficient architecture search. *International Conference on Learning Representations* (pp. 1-13). Vancouver, WA: Carnegie Mellon University. doi:https://doi.org/10.48550/arXiv.1711.00436

Lukac, D., Milic, M., & Nikolic, J. (2018). From artificial intelligence to augmented age an overview. *Zooming Innovation in Consumer Technologies Conference*, (pp. 100-103). doi:https://doi-org.proxy1.ncu.edu/10.1109/ZINC.2018.8448793

Lytras, D., Evaggelos, S., Paris, I., Konstantinos, K., Ioannis, M., & Anastasios, K. (2022). Recording of falls in elderly fallers in Northern Greece and evaluation of aging health-related factors and environmental safety associated with falls: a cross-sectional study. *Occupational Therapy International*, 1-11. doi:10.1155/2022/9292673

Makarenko, O., & Osaulenko, V. (2018). Application of cellular automates in some models of artificial intelligence. *IEEE First International Conference on System Analysis & Intelligent Computing* (pp. 1-4). Kyiv, Kyiv City, Ukraine: Institute of Electrical and Electronics Engineers. doi:10.1109/SAIC.2018.8516837

Mejia, J., Quintero, D., & Builes, J. (2017). Knowledge-based model to support decision-making when choosing between two association data mining techniques. *Revista Lasallista de Investigación, 14*(2), 41-50. doi:https://doi.org/10.22507/rli.v14n2a4

Meta AI. (n.a.). *Tanh Activation*. Retrieved from Papers with Code: https://paperswithcode.com/method/tanh-activation

Miyakawa, T. (2020). No raw data, no science: another possible source of the reproducibility crisis. *Molecular Brain, 13*(1). doi:https://doi.org/10.1186/s13041-020-0552-2

Morris, J. (2008). *Disability research and policy: current perspectives.* Lawrence Erlbaum Associates.

Mullennex, L., & Bachmeier, N. (2023). *Computer Vision on AWS.* Packt Publishing.

Ng, A. (2016). *Machine Learning*. Retrieved from Coursera: https://www.coursera.org/learn/machine-learning

Nguyen, M., Huynh, N., Tran, D., & Ngo, H. (2019). Face recognition applied for smarthome using SoC. *Advanced Computing and Applications* (pp. 165-170). Nha Trang, Vietnam: IEEE. doi:https://doi-org.proxy1.ncu.edu/10.1109/ACOMP.2019.00033

Orhan, A. (2021). How much human-like visual experience do current self-supervised learning algorithms need in order to achieve human-level object recognition? *Neural and Evolutionary Computing*. doi:https://doi.org/10.48550/arXiv.2109.11523

Owen, C. (2017, November 8). *A theorical hands-on introduction to fouculdian discourse analysis*. Retrieved from YouTube: https://www.youtube.com/watch?v=6I6b3ePAZ5M

Oxford. (2022). *Occam's razor*. Retrieved from Lexico: https://www.lexico.com/en/definition/occam's\_razor

Oxford. (2023). *Effectiveness*. Retrieved from Oxford Dictionary: oxfordlearnersdictionaries.com

Ozier, O. (2021). Replication redux: the reproducibility crisis and the case of deworming. *World Bank Research Observer, 36*(1), 101-130. doi:https://doi.org/10.1093/wbro/lkaa005

Parker, R. (1993). Threats to the validity of research. *Rehabilitation Counseling Bulletin, 36*(3), 130-138. Retrieved from https://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=eric&AN=EJ458938&site=eds-live

Phua, K. H. (2021). Ageing in Asia: beyond the Astana declaration towards financing long-term care for all. *International Journal of Health Policy and Management, 10*(1), 32-36. doi:https://doi.org/10.34172/ijhpm.2020.15

Piirainen, K., & Gonzalez, R. (2013). Constructive synergy in design science research: a comparative analysis of design science research and the constructive research approach. *Liiketaloudellinen Aikakauskirja, 3*(4), 206-234. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=bth&AN=95116694&site=eds-live

Qiu, L., Wang, Y., & Rubin, J. (2018). Analyzing the analyzers: FlowDroid/IccTA and AmanDroid,. *ISSTA’18, July 16–21, 2018, Amsterdam, Netherlands*.

Rivera-Landos, E., Khomh, F., & Nikanjam, A. (2021). The challenge of reproducible ML. *International Conference on Software Quality, Reliability and Security (QRS).* *21*, pp. 1079-1088. Hainan Island, China: IEEE. doi:10.1109/QRS54544.2021.00116

Sethi, T., & Kantardzic, M. (2018). Data driven exploratory attacks on black box classifiers in adversarial. *Neurocomputing, 289*, 129-143. doi:10.1016/j.neucom.2018.02.007

Shirazi, B., & Shekhani, S. (2021). Patient’s expectations of privacy and confidentiality in Pakistan. *The Journal of the Pakistan Medical Association, 71*(2A), 537-539. doi:https://doi.org/10.47391/JPMA.888

Silvestrini, R. P., & Sammito, G. (2012). Design of experiments for information technology systems. *Defense AT&L, 41*(5), 30-35. Retrieved from https://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=bth&AN=80409129&site=eds-live

Smaira, L., Carreira, J., Noland, E., Clancy, E., Wu, A., & Zisserman, A. (2020, October 21). A short note on the Kinetics-700-2020 human action dataset. *Computer Vision and Pattern Recognition*. doi:https://doi.org/10.48550/arXiv.2010.10864

Smedley, R. (2019, December 4). *Rob Smedley from Formula 1 talks about using AWS to improve the fan experience*. Retrieved from YouTube: https://youtu.be/eBX7lPk5qmA

Snee, R. (2015). Practical approach to data mining. *Quality Engineering, 27*, 477-487. doi:10.1080/08982112.2015.1065322

Sonmez et al. (2018). Anomaly detection using data mining methods in IT systems: a decision support application. *Sakarya University Journal of Science, 22(4)*, 1109-1123.

Starmer, J. (2017). *What is principal component analysis*. Retrieved from YouTube: https://www.youtube.com/watch?v=HMOI\_lkzW08

Tan, Z. (2021). Ethics events and conditions of possibility. *Business Ethics Quarterly, 31*(1), 106-137. Retrieved from https://search-ebscohost-com.proxy1.ncu.edu/login.aspx?direct=true&db=edb&AN=147839336&site=eds-live

Thomas, R. (2019). The new era of NLP. *Scientific Computing with Python.* Austin, Texas: SciPy. Retrieved from https://youtu.be/KChtdexd5Jo

Tian, Y., Schuemie, M., & Suchard, M. (2018). Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *International Journal of Epidemiology, 47*(6), 2005-2014. doi:https://doi.org/10.1093/ije/dyy120

Ting, W., Chun-Yang, C., Di, G., Xiao-ming, T., & Heng, W. (2014). Clock synchronization in wireless sensor networks: A new model and analysis approach based on networked control perspective. *Mathematical Problems in Engineering Volume 2014, Article ID 731980*, 1-19.

Tun, S., Madanian, S., & Mirza, F. (2021). Internet of things (IoT) applications for elderly care: a reflective review. *Aging Clinical & Experimental Research, 33*(4), 855-867. doi:10.1007/s40520-020-01545-9

Ünal, H., & Başçiftçi, F. (2021). Evolutionary design of neural network architectures: a review of three decades of research. *Artif Intell Rev, 55*, 1723-1802. doi:https://doi.org/10.1007/s10462-021-10049-5

Upchurch, M. (2018). Robots and AI at work: the propects for singularity. *New Technology, 33*(3), 205-218. doi:10.1111/ntwe.12124

US Bureau of Labor Statistics. (2020, May). *Registered nurses*. Retrieved from US Bureau of Labor Statistics: https://www.bls.gov/ooh/healthcare/registered-nurses.htm

Vosshall, P. (2018, November 27). *How AWS minimizes the blast radius of failures*. Retrieved from YouTube: https://youtu.be/swQbA4zub20

Waal, d., & Toit, d. (2011, May/June). Automation of generalized additive neural networks for predictive data mining. *Applied Artificial Intelligence, 25*(5), 380-425. doi:10.1080/08839514.2011.570156

Whitson, G. (2020). *Artificial Intelligence*. Retrieved from Salem Press Encyclopedia of Science: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,sso&db=ers&AN=89250362&authtype=sso&custid=s1229530&site=eds-live&scope=site

Wilensky, U. (2014). *BeeSmart hive finding*. Retrieved from Netlogo: https://ccl.northwestern.edu/netlogo/models/BeeSmartHiveFinding

Zhang, J., Johnstone, M., Le, V., Khan, B., Anwar Hosen, M., Creighton, D., . . . Lynch, M. (2021). Dynamic time warp-based clustering: Application of machine learning algorithms to simulation input modelling. *Expert Systems with Applications, 186*. doi:https://doi.org/10.1016/j.eswa.2021.115684

Zhao, W. (2014). *Building dependable distributed systems.* John Wiley & Sons, Incorporated.

Zhu, X., Zhu, J., Li, H., Wang, X., Li, H., Wang, X., & Dai, J. (2021). Uni-perceiver: pre-training unified architecture for generic perception for zero-shot and few-shot tasks. *Computer Vision and Pattern Recognition*. doi:https://doi.org/10.48550/arXiv.2112.01522