Chapter 2: Literature Review

Nate Bachmeier

DIS-9902: Dissertation Proposal

April 24, 2022

Northcentral University

# Chapter 2: Literature Review

The problem to be addressed in this study is implementing a quality assurance process for an autonomous assistant for elderly and special needs care. This research theorizes that using 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, Madanian, & Mirza, 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 down innovation and restrict the value researchers can contribute to the body of knowledge.

This constructive research design study aims to propose a research process that divorces privacy and safety concerns from investigating autonomous assistants in elderly and special needs care. It aims to deliver this capability by utilizing humanoid constructs within a realistic physics simulation process. Next, positioning virtual cameras, instruments, and devices within the virtual world enables researchers to collect their experimentation data. Lastly, the automation can modify the environment using programmable interfaces such as raising the alarm or applying other mitigations.

## Chapter overview

*Provide an overview of the subheadings in the literature that will discuss. This section should be around a 1-page and help frame what the reader will learn from this 50-page document. Otherwise, it risks being very confusing and not meeting their expectations.*

## Search Method and Resources

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 Institute of Electrical and Electronics Engineers (IEEE), Association for Computer Machinery (ACM), Springer Publishing, John Wiley & Sons ProQuest, among others.

A breath-first search scanned for surveys, challenges, and opportunities on the constructive research project’s core concepts (see Table 1). 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 document extracted themes. Those sorted themes are available in the proceeding conceptual frame section, which attempts to present each topic’s current state and direction from Table 1.

Table 1: 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 |

## Conceptual Framework

A conceptual framework is a blueprint that communicates a natural progression of the phenomenon to be studied (Dickson, Emad, & Adu-Agyum, 2018). It is essential for quality research as it outlines a methodical structure of definitions, concepts, and relationships.

There are four approaches to studying a business use case or phenomena (see Table 2). 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 2: 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) criticizes 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 focusing on 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 delivering 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 come with 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) is a process for extracting data from image sources. Next, human activity recognition (HAR) processes must classify that data into distinct actions and behaviors, such as the person is 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 needs to 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 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 them with the business challenge of improving elderly care and special needs. It also uses generalizable virtual camera instruments for CV and HAR experimentation with humanoid agents. Aspects of this framework appear in other publications (Gu et al., 2021; Banjarey et al., 2021). However, utilizing humanoid subjects with HAR and CV is not a mainstream topic 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). There are several advantages to wearable sensors, such as they follow the patient within the world. Researchers have also demonstrated applying these sensors to predicting diverse action spaces (Nugroho et al., 2018). However, CV-based agents can extract more context from those same behaviors. For instance, a biosensor might predict that a patient is eating, but not the food type. Instead of directly competing technologies, future solutions must integrate these heterogeneous sources.

Second, 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, it might be more challenging to train models on this real-world basis versus 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.

## Challenges and opportunities for care providers

*This section is a placeholder for compiling notes from the Industry state section. It attempts to frame the business environment and limitations that create the need for additional research.*

## 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 gave life. The sprouting of ideas came from mathematics, biology, and computer science before eventually producing modern artificial intelligence. While these different domains have unique perspectives, they collectively land in four categories of intelligent systems (Lukac, Milic, & Nikolic, 2018). The first divide 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, specifically 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, lay the foundation that became machine learning. Researchers use these tools to build statistical models that represent a situation. For instance, if a customer purchases bread, what else could you recommend? 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, Milic, & Nikolic, 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 to 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, Rust, & Maksimovic, 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, emulate interpersonal experiences, and express 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 that considers these characteristics and their weighted importance. Meanwhile, another company needs to classify handwritten digits, which requires 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, Liu, & Fogel, 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 called neural dynamics, embodiment, and awareness (Ballard & Zhang, 2021).

### 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 applying 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 that have 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 since 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 the visual information.

## Recognizing human activities

One critical application of CV 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 (Gu et al., 2021; Banjarey 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)

A preceding section examined the biological constructs that enable primates’ vision and nature’s solution.

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 that a fully trained model requires at least ten observations per parameter (Snee, 2015). Depending on the connectivity configuration, this 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 1). 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 1: Network Structure

Diagram

Description automatically generated

### Recurrent Nural Networks (RNN)

### Generative Adversarial Networks (GAN)

### Graph Convolutional Neural Networks (G-CNN)

## How does human activity recognition (HAR) work

HAR is a critical component for health care

### Traditional methods

*There are currently no documents for this section. Need it?*

### Deep learning methods

* Learning models for HAR (2021)
* Survey on HAR using Sensors (2021)
* Survey on DNN for HAR (2021)

### Graph CNN methods

## How does ML model training work

*Include a description of the problem and briefly describe the approaches.*

### Hardware acceleration

### Continuous Learning

### Optimization Patterns

## How do dynamic environment simulations work

### Methods

### Unity-based

# References

Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE, 76*(8), 966-1005. doi:10.1109/5.5968

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

Banjarey, K., Prakash, S., & Kumar, D. (2021). A survey on human activity recognition using sensors and deep learning methods. *5th International Conference on Computing Methodologies and Communication* (pp. 1610-1617). Erode, India: IEEE. doi:https://doi.org/10.1109/ICCMC51019.2021.9418255

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

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

Gu, F., Chung, M., Chignell, M., Valaee, S., B, Z., & Liu, X. (2021). A survey on deep learning for human activity recognition. *ACM Computing Surveys, 8*, 1-34. doi:https://doi.org/10.1145/3472290

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75-105. doi:10.2307/25148625

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

Iivari, J. (2020). A critical look at theories in design science research. *Journal of the Association for Information Systems, 21*(3), 502-519. doi:doi:10.17705/1jais.00610

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

Marr, D. (n.d.). *Vision a computational investigation into the human representation and processing of visual information.*

Nugroho, H., Harmanto, D., & Al-Absi, H. (2018). On the Development of Smart Home Care: Application of Deep Learning for Pain Detection. *Biomedical Engineering and Sciences* (pp. 612-616). IEEE. doi:10.1109/IECBES.2018.8626710

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

Prapas, I., Derakshan, B., Mahirji, A., & Markl, V. (2021). Continuous Training and Deployment of Deep Learning Models. *Datenbank-Spektrum, 21*(3), 203-213. doi:https://doi.org/10.1007/s13222-021-00386-8

Right Vision Media. (2021, March 5). Global data center accelerators market trajectory & analytics 2020-2027. *Gale General OneFile*, NA. Retrieved from https://ink.gale.com/apps/doc/A653973707/ITOF

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

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

Zeller, A. (2014). *What makes useful research in software engineering*. Retrieved from YouTube: https://youtu.be/4MbixFVWwck