The Ontology of Longevity

Aaron Jin aaronjin@stanford.edu

Susan Lee leesusan@stanford.edu

Kevina Wang kevinaw@stanford.edu

Abstract

In 2015, for the first time since 1993, life expectancy in the United States year-over-year[1]*. Propelled by disturbing trends in societal mortality, the pursuit for longevity has become a global phenomenon, with a market valued at* 25.1 *billion in 2020, projected to reach* 44.2 *billion by 2030* [2]*. In the burgeoning field of longevity, understanding the complex interplay between genetic, environmental, and social factors is crucial for advancing and leveraging research.This project integrates three key components to create an intuitive platform that recommends individual's tailored lifestyle choices supporting longevity:*

1. *An ontology encompassing the broad spectrum of longevity factors and their relationships.*
2. *A backend that leverages NLP techniques and OpenAI's GPT model, trained on the project's ontology*
3. *A frontend that enables individuals to interact with the recommendation system*

The ontology was created using Protege employing Web Ontology Language, OWL. The backend utilizes semantic embeddings, powered by a pre-trained transformer model. OpenAI's GPT model intake users' responses and provide personalized recommendations. The frontend of this project was developed using React and deployed via Vercel.

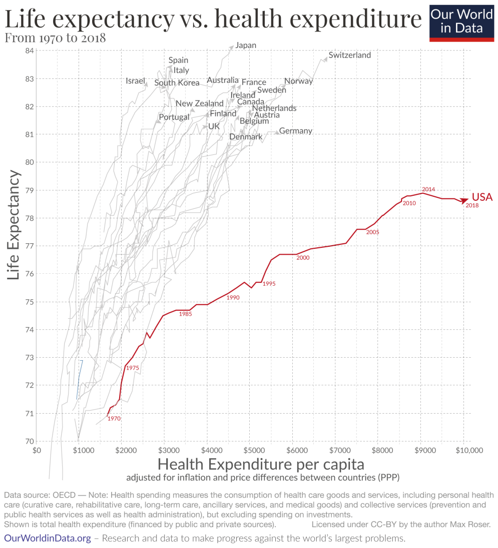
Generated suggestions from user-prompts effectively provided tailored recommendations, addressing physiological, psychological, and environmental factors. Looking ahead, refining the project's ontology by including additional factors, such as cultural influences, could enhance its precision and relevance. Furthermore, optimizing suggestions to align with users' practical realities by soliciting contextual information from users could ensure the recommendations are not only personalized, but also realistically actionable.

1 Background and Motivation

1.1 Trends in Lifespan in the United States

Year-over-year life expectancy in the United States has been rising with rare exceptions since the 1900s. In 1900, the average life expectancy in the United states was 47 years. In 1950, average life expectancy was 68 years, and in 2019, 79 years. However, in 2020 and 2021, life expectancy from fell to 77 and 76 years, "representing the largest decrease over a two-year span since the 1920s." [3] Notably at this time, decreases in life expectancy are disproportionately observed in American Indian and Alaska Native populations, with decreases in life expectancy between 2020 and 2022 at 4.0, 4.2, and 6.6 years respectively for Black, Hispanic, and American Indian/Alaska Native populations. This is contrasted with 2.4 and 2.1 years for white and Asian Americans populations. While the CDC attributes "75% of the decline from 2019 to 2020 and 50% of the decline from 2020 to 2021 to COVID," the decline remains paramount. Despite decreases in deaths from leading causes of death such as chronic lung disease, pneumonia, influenza, and Alzheimer's, net declines in life expectancy were driven by increased incidence of heart disease, liver disease, drug overdose, and suicide [3].

Life expectancy in the United States is considerably lower than other rich countries, despite spending more on healthcare [5].



The diminished life expectancy in the United States relative to other affluent countries is linked to a plethora of factors leading to higher mortality rates, including tobacco use, obesity-related complications, homicides, opioid misuse, suicides, and road incidents. Obesity is a prevalent issue, with 70% of the American population being overweight and 36% falling into the obesity category, conditions which are risk factors for diseases such as cardiac ailments, diabetes, certain forms of cancer, and strokes. The opioid crisis in the United States is severe, with death rates from overdoses being three times higher than in any other high-income nation. Despite a global trend of declining suicide rates, the United States has experienced a slight increase, with suicides, notably with those involving firearms. Furthermore, the disparity in life expectancy is also exacerbated by the absence of universal health care coverage, which is found in all other developed nations.

1.2 Private Sector Funding

Alarming recent patterns in lifespan have led to a proliferation of interest in longevity in the private sector. Research and investment in longevity therapeutics, biotechnology, and regenerative medicine have surged as stakeholders seek solutions to reverse declining life expectancy trends and address the underlying factors contributing to premature mortality. Venture capital funding for longevity is led by firms including LongeVC ($35 million fund), LifeX VC ($100 million fund), and Longevitytech.fund ($150 million fund). Notable start-ups include Chroma Medicine (epigenetic editing therapeutics), Precision Biosciences (ARCUS genome editing), and Tune Therapeutics (fine-tuning activities in cells) $150 million)[4]. Big Tech has also indicated interest in longevity, with Bezos, executive chairman of Amazon, having invested in Altos Labs, a start-up whose mission is to restore cell health and resilience through cell rejuvenation to reverse disease, injury, and the disabilities that can occur throughout life [5]. Established biotechnology companies, such as Calico Labs, have also put their foot in the door, with Calico Labs researching aging reversal in cells.

1.3 Factors Influencing Longevity

1.3.1 Biological

There are a variety of biological factors that are known to influence longevity and are hot topics of research. Research to identify specific genes that control aging have continued to elude researchers, and although genes in other specimens such as the fruit fly have been identified, identifying singular genes in the human genome has remained relatively inconclusive. However, the E2 allele of the APOE gene in humans has been found in centenarians and is a strong candidate for a gene that could influence longevity. In addition, genes associated with genetic disorders, such as the WRN gene in Werner syndrome (rare autosomal recessive adult-onset progeroid with early manifestations of aging such as hair loss, skin atrophy, premature heart disease, and various tumors) have also been discovered. In addition to genes, cellular mechanisms involved with aging have also been discovered, such as the 12 hallmarks of aging, which includes factors like stem cell exhaustion and cellular senescence. These discoveries have led to various theories in how aging occurs: In Kirkwood's paper "Why Can't We Live Forever?", Kirkwood proposes the disposable soma theory, which states that "senescence is the price paid for sexual reproduction." In other words, since our body balances between cellular aging and repair versus reproduction, ultimately aging occurs as cellular damage accumulates over time. Lastly, another topics of interest within biological factors of longevity is the compression of morbidity, proposed by James Fries, which states the goal to increase the age of which chronic disease symptoms first appear to increase our health span, not just our life span.

1.3.2 Environmental

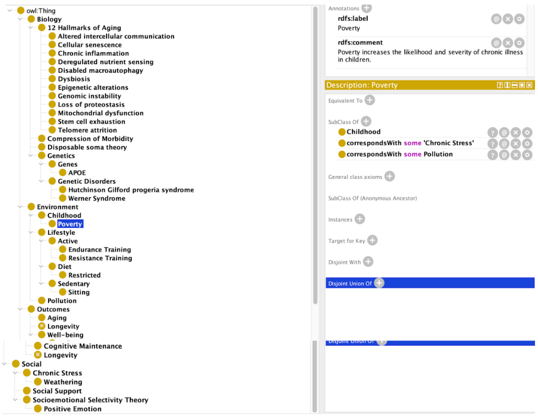
Twin studies have shown that biological factors are not the only factors influencing longevity. In fact, twin studies have revealed that most of the variance in lifespan attributes to non-shared environmental factors. For example, childhood conditions have been shown to have significant contributions to adult health. There is growing evidence that pollution can generate changes in neurodevelopmental and immunological processes. In addition, early exposure to chronic stress has also been suspected to lead to altered development of brain structures. Failed access to effective interventions also plays a role in transforming poverty's influences from early life to adulthood. Finally, one's lifestyle also plays a significant role in longevity; less physical activity is associated with a higher ratio of mortality, and somewhat concerningly, more people are starting to report less physical activity over the years.

1.3.3 Social

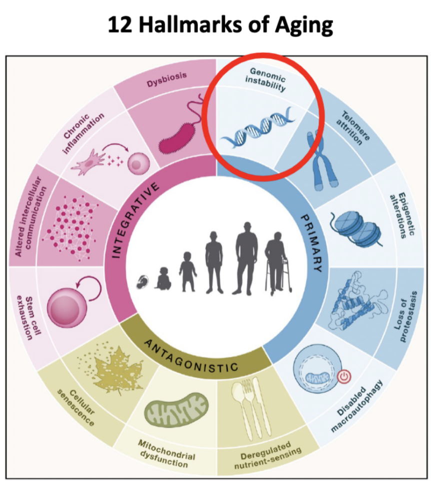
Lastly, there are a variety of social factors that have also been studied in affecting longevity. Weathering is a process where the day-in, day-out effort of trying to be equal wears bodies down from chronic stress (which happens more often in marginalized populations). Weathering and chronic stress are associated with earlier onsets of hypertension, diabetes, and strokes, and high cortisol levels have also been seen passed down from pregnant mothers to their developing children. Another social aspect in regards to longevity is the socioemotional selectivity theory, developed by Laura Carstensen. This theory proposes that an approach of endings (whether from aging, relocation, or illness), elicits motivational changes in which emotionally meaningful goals are prioritized over exploration. Overall, as goal priorities change with age, these goals direct cognitive resources, and these shifts can influence top-down cognitive processing that can affect people's thinking, lifestyles, and more. Finally, living a fulfilled life, whether from a close-knit circle or a strengthened sense of purpose in life, can lead to longer, healthier lives. Finally, education is another social factor that affects and could be affected by longevity. While our current education system mainly focuses on education in early years, as lifespans continue to increase, a more continuous education is expected to not only lead to more fulfillment in later lives, which is linked with happiness and health in elderly populations, but also create more inter-generational work-forces and families.

2 Methods

2.1 Ontology



This project's proposed ontology categorizes factors influencing longevity into biological, environmental, and social factors. Central to the biological factors are the "12 Hallmarks of Aging," which include the biological mechanisms thought to contribute to the aging process.



The biological section also highlights specific genes and disorders such as APOE, Hutchinson Gilford progeria syndrome, and Werner Syndrome. Environmental influences categorized into childhood, lifestyle, and pollution factors. Lifestyles, including active (endurance and resistance training) and sedentary behaviors, are categorized, alongside diet and pollution. Finally, the ontology addresses social aspects, such as chronic stress and social support, highlighting their role in the aging process. We sourced much of the concepts and data for this ontology from literature and domain experts (professors), and also were inspired by the class PSYCH 102 (Longevity), that provided the broad framework our ontology has. After breaking down the ontology into broad categories, we also created several relationships that could help users understand that while we separated our ontology into categories, these categories are all naturally intertwined, and they all affect each other. For example, we created the properties "combats", "contributesTo", and "correspondsWith" to connect much of our classes together. To provide several examples, chronic stress contributesTo some aging, social support combats some poverty, and diet corresponds with some cognitive maintenance. We also added an Outcomes class in our ontology to represent the different areas of longevity our ontology covers. While we have material that discusses well-being and increasing lifespan, such as active lifestyles, positive emotion, and the APOE gene, we also ensured that our ontology would be multi-dimensional in also discussing what harms longevity, including genetic disorders such as Hutchinson Gilford progeria syndrome, weathering, and poverty.

While we initially divided the Outcomes class into two categories, just aging and well-being, we thought it would be more fitting to also include longevity as a class itself, as there is some nuance between well-being and longevity (for example, the APOE gene contributes to longevity but doesn't necessarily contribute to a person's well-being; someone could still be unhappy or have chronic stress but still contain a gene that naturally extends their lifespan). To address this nuance, we defined longevity specifically as equivalent to "Well-being and (combats some Aging), so that longevity doesn't necessarily equal well-being but doesn't necessarily exclude it, and also must address aging in some way.

Overall, our ontology includes a mix of theories, ideas, tangible factors, and actions that affect longevity in some way, or are related to the field in some way. While it is difficult to truly encompass all of the ideas and players involved in longevity (whether in research, economoics, social policy, or business), we believe that this ontology does its best to represent the broader relationships and ideas that can be found when understanding longevity.

2.2 Problem-Solving Methods

In our project, we developed problem-solving methods tailored to operate on our ontology, integrating key factors related to lifestyle, aging, and longevity. Our approach integrates both rule-based logic and machine learning (ML) techniques to analyze and interpret the ontology, allowing us to generate personalized recommendations and insights based on user inputs. This methodology not only enhances the user's interaction with the system but also ensures that the generated outcomes are both relevant and scientifically grounded, thereby facilitating informed decision-making in the context of health and well-being. To take a look at our code, click here.

2.2.1 Backend

In the development of the backend for our project, we employed a stack of technologies and frameworks chosen for their robustness, scalability, and ability to handle complex natural language processing (NLP) tasks.

Our backend architecutre is fundamentally built on Python, which facilitates the integration of our chosen libraries and frameworks essential for our NLP and AI-driven tasks.

We leverage the OpenAI API, which provides access to powerful AI models including GPT-4. This API is crucial for generating intelligent, contextually relevant responses based on user inputs. Additionally, the API's robustness allows for the processing of complex queries and the generation of nuanced text responses and recommendations.

For the processing and understanding of natural language, we integrate the Transformers library, a suite of ML models for our NLP tasks. More specifically, we utilize the BERT (Bidirectional Encoder Representations from Transformers) model to generate embeddings for text inputs. These embeddings are high-dimensional vectors that capture the semantic properties of the user's inputted text (as well as the text found within the ontology itself), enabling our system to understand natural language effectively and accurately.

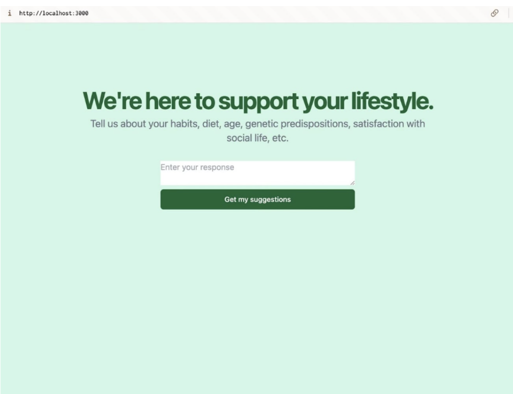
Our backend also uses PyTorch, an ML library, to facilitate the manipulation and computation of tensors, which are essential for the operation of the Transformers models. We found that this library's dynamic computation graph and efficient memory usage made it particularly useful for the real-time processing of text data and the generation of embeddings.

To manage and interpret semantic web data (the data found within our ontology), we use RDFLib, a Python library for working iwth RDF data. This library allows for our code to parse and manipulate RDF data, which is structured in a way that facilitates semantic queries and the extraction of meaningful data from the dataset.

Lastly, for calculating the similarity between the user inputs and the ontology data (stored in RDF format), we employ the cosine similarity metric from the Sklearn library. This metric measures the cosine of the angle between two vectors in a multi-dimensional space, essentially meaning that it provides a quantifiable measure of similarity. This is used to identify the most relevant ontology nodes, paired with the nodes' descriptions, related to the user's query.

Our backend system combines these various technologies and frameworks to handle the different aspects of NLP and AI-driven interactions for our project. This blend of technologies ensures that our project is both efficient and effective in processing and effective in processing and responding to user inputs, given our ontology data.

2.2.2 Frontend



For the frontend of our project, we crafted an interface that bridges the user's interaction with the underlying backend ontology and NLP functionalities.

The cornerstone of our frontend architecture is Next.js, a React framework that allows us to facilitate server-side rendering, static site generation, and the creation of a web application like ours. We chose Next.js for its comprehensive ecosystem and developer-friendly features, allowing us to accelerate our overall development process. Through this framework, we developed a responsive and interactive web page where users are prompted to input details regarding their lifestyle and any known risk factors. The design focuses on ensuring that users can easily engage with the system.

The frontend's layout and styling was built through semantic HTML and Tailwind CSS. Semantic HTML is employed to structure the web page's content meaningfully. On the other hand, we utilized Tailwind CSS for its modular approach to styling.

Our frontend system is designed with a focus on user experience by leveraging the strengths of using TypeScript, Tailwind CSS, as well as Next.js and React so that users can interact with our ontology on lifestyle, aging, and longevity. This design ensures that users are in an intuitive environment to input their lifestyle data and genetic risk factors.

2.3 Evaluation

To evaluate the performance and effectiveness of our project, we employ a combination of objectivistic and subjectivistic approaches. Our evaluation process is designed to assess the system's ability to provide accurate, relevant, and actionable recommendations to users based on their input and our underlying ontology.

2.3.1 Objectivistic Evaluation

For the objectivistic evaluation, we focus on quantitative metrics and measure our system's performance in terms of accuracy and relevance of its generated recommendations. As discussed in class, we employ a combination of automated metrics and manual analysis, somewhat similar to the MYCIN evaluation approach, to assess the quality of these recommendations.

First, we collect a diverse set of user inputs that cover various lifestyle factors, genetic predispositions, and environmental influences. For each input set, our system generates a set of recommendations. We then compare these recommendations against a predefined set of guidelines as well as our collection of longevity research's best practices, both of which are derived from the ontology and by domain experts in the realm of longevity research itself. This comparison allows us to calculate the precision and recall of our system's recommendations.

Additionally, we assess the semantic similarity between the user inputs and the generated recommendations using cosine similarity metrics. This helps to ensure that the recommendations are contextually relevant to the user's specific input, rather than being generic or unrelated.

To further validate the objectivistic evaluation, we employ cross-validation techniques (like k-fold cross-validation) to assess how our systems performs across different subsets of the user input data. This helps to mitigate potential biases and ensures the robustness of our evaluation results.

2.3.2 Subjectivistic Evaluation

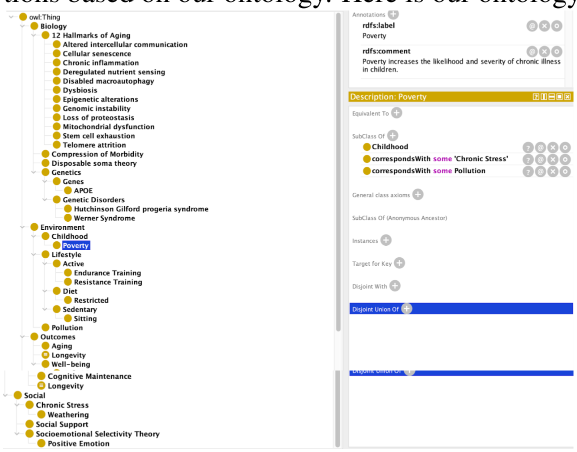
To complement the objectivistic evaluation, we also conduct a subjectivistic evaluation that focuses on the user experience and the perceived value of the system's recommendation. This form of evaluation is a bit harder to measure as it is not quantitative. We decided to employ a focused-methods approach, using user interviews to gather qualitative insights into how users interact with out system and perceive its impact on their decision-making processes related to longevity.

To do so, we recruit a set of users, representing various lifestyles and health conditions, to participate in this evaluation. We ask users to interact with our project, inputting their personal data in the form of a text response and receiving recommendations. Through semi-structured interviews, we collect feedback on the clarity and relevance of these recommendations as well as the users' overall satisfaction of said recommendations.

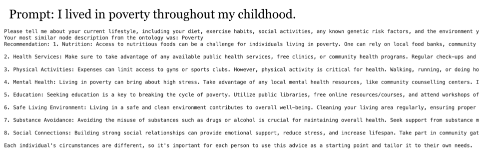
By combining the objectivistic and subjectivistic evaluation approaches, we aim to gain a more comprehensive understanding of our system as well as its performance, user acceptance, and potential impact on users' longevity-related decisions and behaviors.

3 Results

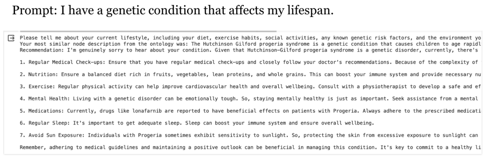
To summarize much of what has been discussed in the methods sections, we created an ontology, trained a model based on our ontology, and created a user-interface to query longevity-related questions based on our ontology. Here is our ontology:



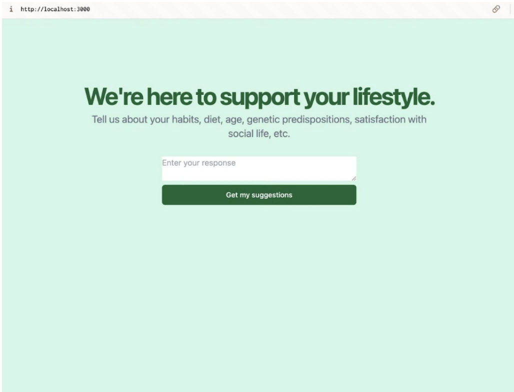
Here are some sample results based on various user queries. The first query is "I lived in poverty.":



Here are the results from another sample query, "I have a genetic condition.":



The final product, the user interface, looks like the following:



Based on our evaluations, we continued to fine-tune our models and add more descriptive labels in our ontology to ensure we could capture the full meaning of classes in our ontology (many of which can't be captured just by relationships alone. For example, the disposable soma theory required a longer description for the model to produce more accurate results regarding the theory if it was ever queried).

4 Discussion and Future Work

This project aims to develop a customized, lifestyle recommendation system built upon a holistic model of factors influencing longevity. This project built an ontology of factors influecning longevity and employed this ontology in a backend using NLP and AI-techniques to create a dynamic framework capable of generating personalized lifestyle recommendations aimed at promoting longevity.

The primary advantages of this project's methodoly lies in its comprehensive view of longevity. The ontology meticulously captures a spectrum of factors affecting longevity as well as how they interplay, via biological, environmental, and social factors and relationships "combats", "contributesTo", "correspondWith". Another notable advantage with this project methodology is its provision of a tangible interface for users to engage with the ontology. Furthermore, this project outperforms existing longevity recommendation capabilities by large language models (LLMs) by agumenting LLMs with validated ontology data. This approach mitigates the risk of hallucinations, thereby enhancing the accuracy and reliability of the lifestyle recommendations provided as compared to general LLMs.

However, this project faces signficant challenges in terms of personalization. The current methodology fails to tailor recommendations based on the unique contexts of users, potentially leading to suggestions that are not feasibly actionable. Furthermore, relationships between factors are surface-level; the magnitude at which one factor "contributesTo", "correspondsWith", or "combats" another is generalized, rather than quantified. Moreover, translating the ontology into a text-based input interpretable by a LLM exposes this model to erroneous interpretations; ontologies and text and inherently different data structures and attempting to translate one into another leads to challenges maintaining the precise semantics and relatioships defined in an ontology, resulting in misinterpretations or oversimplifications when processed by the LLM.

Immediate objectives for this project would be enhacing personalization to respect user context and resource bandwidths, quantifying relationships within the ontology, and refining to translation process from ontology to LLM input. Subsequent priorities include developing code to protect user data privacy, integrating more factors into the ontology to immprove comprehensiveness, and expanding the project to support multiple languages. These steps are critical to enhancing system accessibility, making long and health lives available to a larger, more global audience.

5 Division of Labor

All labor on this project was divided equally, and all members worked equally on this project. While everyone had a say and part in each aspect of the project, Susan focused on research and building the ontology, Kevina focused on the frontend, and Aaron focused on the backend.

6 Authorization to Share

We do not authorize sharing this project.

7 Appendix

1. Bosworth, B., & Zhang, K. (n.d.). *Life expectancy and inequality in life expectancy in the United States*. Stanford Institute for Economic Policy Research (SIEPR). Retrieved from https://siepr.stanford.edu/publications/policy-brief/life-expectancy-and-inequality\-life-expectancy-united-states

2. Strategy&. (n.d.). *The potential of longevity therapeutics*. Retrieved from https://www.strategyand.pwc.com/de/en/industries/pharma-life-science/longevity-therapeutics.html

3. McCarthy, C. (2022, October 20). *Why life expectancy in the US is falling*. Harvard Health Blog. Retrieved from https://www.health.harvard.edu/blog/why-life-expectancy-in-the-us\-is-falling-202210202835

4. Majic, J. (2023, November 9). *Venture capital giants in the billion-dollar quest for longevity breakthroughs*. Forbes. Retrieved from https://www.forbes.com/sites/josipamajic/2023/11/09/venture-capital-giants-in-the\-billion-dollar-quest-for\-longevity-breakthroughs/?sh=21212c011265

5. Altos Labs. (n.d.). Retrieved from https://www.altoslabs.com/

6. Fries, J. F. (2005). *The compression of morbidity. 1983*. The Milbank Quarterly, 83(4), 801–823. https://doi.org/10.1111/j.1468-0009.2005.00401.x

7. Finch, C. E., & Tanzi, R. E. (1997). *Genetics of aging*. Science (New York, N.Y.), 278(5337), 407–411. https://doi.org/10.1126/science.278.5337.407

8. Kirkwood, T. (2024, February 20). *Will people ever live forever?*. Scientific American. Retrieved from https://www.scientificamerican.com/article/will-people-ever-live-forever/

9. Johnson, A., & Gomez, C. (2023, October 17). *Stress is weathering our bodies from the inside out*. Washington Post. Retrieved from https://www.washingtonpost.com/health/interactive/2023/stress-chronic-illness-aging/

10. Carstensen, L. L. (2021). *Socioemotional Selectivity Theory: The Role of Perceived Endings in Human Motivation*. The Gerontologist, 61(8), 1188–1196. https://doi.org/10.1093/geront/gnab116

11. Wise, P. H. (2016). *Child Poverty and the Promise of Human Capacity: Childhood

as a Foundation for Healthy Aging. Academic Pediatrics, 16(3 Suppl), S37–S45. https://doi.org/10.1016/j.acap.2016.01.014