Seth Howell

CS 402: Independent Study

Dr. Derrick Tate

5/13/2021

The Intersection of Psychometrics and Artificial Intelligence

Intro

Over the course of this semester, I have performed research to prepare for my capstone. My research covered some of the history and theory of psychometrics as well as artificial intelligence. I have developed a goal to use computer programming and data processing tools within artificial intelligence to perform personality assessment.

Personality Psychology

Personality psychology is a popular field globally. The Oxford English Dictionary defines personality as "That quality or assemblage of qualities which make a person what he is, as distinct from other people; a distinctive personal or individual character, especially of a marked or noticeable kind."[1] Psychometrics is the field of psychology that seeks to measure personality through assessments.[2] A $2 billion industry has developed for personality assessment, being utilized by corporations, schools, churches, and by individuals.[3] The most popular assessments fall into the *type* category.[4] A participant fills out a lengthy questionnaire and is then given a personality type based off his or her answers.

The practice of restricting complex humans to singular personality types has triggered significant pushback. Psychologist Benjamin Hardy in his book *Personality Isn't Permanent* highlights five dominant views on personality:

1. Personality can be categorized into "types."
2. Personality is innate and fixed.
3. Personality comes from your past.
4. Personality must be discovered.
5. Personality is your true and "authentic" self.

---

[1] *Oxford English Dictionary*. https://www.ocf.berkeley.edu/~jfkihlstrom/PersonalityWeb/Ch1Personality.htm. Accessed 5/12/2021.

[2] https://www.understandmyself.com/. Accessed 5/17/2021.

[3] M. Emre, *The Personality Brokers: The Strange History of Myers-Briggs and the Birth of Personality Testing*, New York: Doubleday (2018).

[4] Three such type assessments are the Myers Briggs Type Indicator, DISC, and Keirsey Temperament Sorter.

He claims that these "pervasive and destructive myths… lead people to adopt a narrow and fixed mindset about themselves."[5] He proposes rather that personality is a by-product of one's decisions in life. He points to world changers like Gandhi whose personalities are shaped through their commitments to a larger vision, not vice versa.[6] Hardy's approach offer's people an open canvas to shape who they want to become rather than restricting them to a *type*. Thus, my capstone will seek to utilize the best parts of psychometrics while also refraining from placing people within boxes.

Artificial Intelligence

AI is an extremely broad field backed by nearly a century of development. Experts classify AI into four major categories:[7]

1. Systems that think like humans.
2. Systems that act like humans.
3. Systems that think rationally.
4. Systems that action rationally.

At the heart of these categories is the process of taking a system (computer) and giving it the capabilities to think and act. According to Yale, one ultimate goal of AI research is "to understand how people are possible—i.e., how it is that an intelligent system can thrive in the real world," to "understand how the human brain extracts information from its environment, and uses it to guide behavior."[8]

AI has advanced greatly toward achieving this goal. In the scope of my capstone, I aim to produce/compile software that is able to extract information from computer users and make accurate assessments about the users accordingly. The ubiquitous practice of assessing personality via self-assessment questionnaires has many flaws. The questions (normally statements), which provide the participant with a spectrum of options (known as a *Likert* scale), are often hard to gauge and produce inconsistent results. For example, the question might be "You rarely do something out of just sheer curiosity," and the Likert scale may be from 1: "strongly disagree" to 5: "strongly agree". A journalist points out by contrast the power of computers to assess users: "your browser already knows what kind of person you are a thousand times more intimately than any test will ever reveal."[9] In theory, AI could be used to produce consistent and accurate personality results for computer users. I hope to prove this in my capstone.

Background

---

[5] B. Hardy, *Personality Isn't Permanent*, New York: Portfolio (2020). Pg. 25.
[6] Ibid, 53.
[7] S. Russell & P. Norvig, *Artificial Intelligence: A Modern Approach*, 2002. https://plato.stanford.edu/entries/artificial-intelligence. Accessed 5/12/2021.
[8] Yale, Computer Science. https://cpsc.yale.edu/research/artificial-intelligence. Accessed 5/12/2021.
[9] https://www.newyorker.com/magazine/2018/09/10/what-personality-tests-really-deliver. Accessed 5/13/2021.

1. Attention Space: Personality Assessment Models

Psychologists have officially been being assessing personality via questionnaires for a century now. Henry Rorschach developed a test in 1921 to evaluate thought disorder and pinpoint mental illness. Participants would view images and then answer questions about what they saw.[10] At this time Sigmund Freud and Carl Jung were making significant contributions to personality psychology. Jung developed a theory of psychological types, the central two being an extraverted attitude and an introverted attitude. While no one operates as exclusively an extravert or introvert, these types frame how one experiences the world. He then proposed four functions which people use to orient themselves to their outer and inner world. Each person primarily uses either thinking, feeling, sensation, or intuition.[11] From the 1920s through the 1940s Katherine Briggs studied Jung's theory and developed four major dichotomies:

1. Extravert (E) vs. Introvert (I)
2. Senser (S) vs. Intuitive (N)
3. Thinker (T) vs. Feeler (F)
4. Judger (J) vs. Perceiver (P)

From combinations of these eight letters are derived sixteen personality types, each person being characterized by the combination of four (e.g., ESFP). Since then, the official Myers Briggs Type Indicator (MBTI), an introspective self-report questionnaire measuring preferences among the four dichotomies, has become the dominant design for personality assessment. MBTI and its newer spinoff, the Keirsey Temperament Sorter are both prominent among corporations, the former being utilized by about 88% of Fortune 500 companies.[12]

A different family of personality assessments has also gained dominance over the past four decades for using more credible science to measure personality. These assessments are concerned with *traits* rather than types and seek to place people on a personality spectrum rather than within a category. In the nineteenth century, psychologist Francis Galton developed the lexical hypothesis, which assumes that people encode in their daily language the individual differences that are important in daily encounters. In the 1930s Gordon Allport and Henry Odbert adopted this hypothesis and identified 17,953 unique terms to describe personality/behavior from a 400,000-word dictionary. About 4,500 of these terms were classified as describing stable personality traits/dispositions.[13] In the 1940s Raymond Cattell analyzed this list and reduced it to about 170 items. Through testing and factor analysis he was able to recognize sixteen global factors which can be measured to assess personality.[14] Since then psychologists have narrowed this list to five

---

[10] https://openpsychometrics.org/tests/HEMCR. Accessed 5/14/2021.

[11] https://philosophy.lander.edu/ethics/jung.html. Accessed 5/14/2021.

[12] https://www.themyersbriggs.com/en-US/Products-and-Services/Myers-Brigg. Accessed 5/14/2021.

[13] G. W. Allport & H. S. Odbert, *Trait-names: A Psycho-lexical Study*, Albany, NY: Psychological Review Company (1936).

[14] H. B. Cattell, "The 16PF: Personality In Depth," Champaign, IL: Institute for Personality and Ability Testing, Inc (1989).

traits: extroversion, agreeableness, openness, conscientiousness, and neuroticism. Each person contains some percentage of these traits and certain conclusions can be drawn according to these percentages (e.g., high conscientiousness is a good indicator for long-term success). The most prominent assessments for measuring these five factors are Cattell's Sixteen Personality Factors, Big Five, NEO-PI-R, and HEXACO.

There are also relatives of the personality type and trait models that have become prominent, such as career interest assessments and value assessments. Some prominent examples of these are the Strong Interest Inventory, The Motivational Appraisal of Personal Potential (MAPP) assessment, and the Motives, Values, Preferences Inventory (MVPI).

Together, all the aforementioned testing models serve a significant role in the field of psychometrics. Here is a map of prominent personality assessments over the past century:

Rorshach	1921.	Big Five	1980s.
MMPI	1939.	Enneagram	1980s.
16PF	1949.	NEO-PI-R	1992.
BM	1951.	SVS	1992.
DISC	1956.	MAPP	1995.
CPI	1956.	SF	1999.
MBTI	1962.	HEXACO	2000.
SII	1972.	MVPI	2010.
EPQ	1975.	High5	2010s.
KTS	1978.	Personabilities	2020.

KEY
Scientist / Thinker
Folk
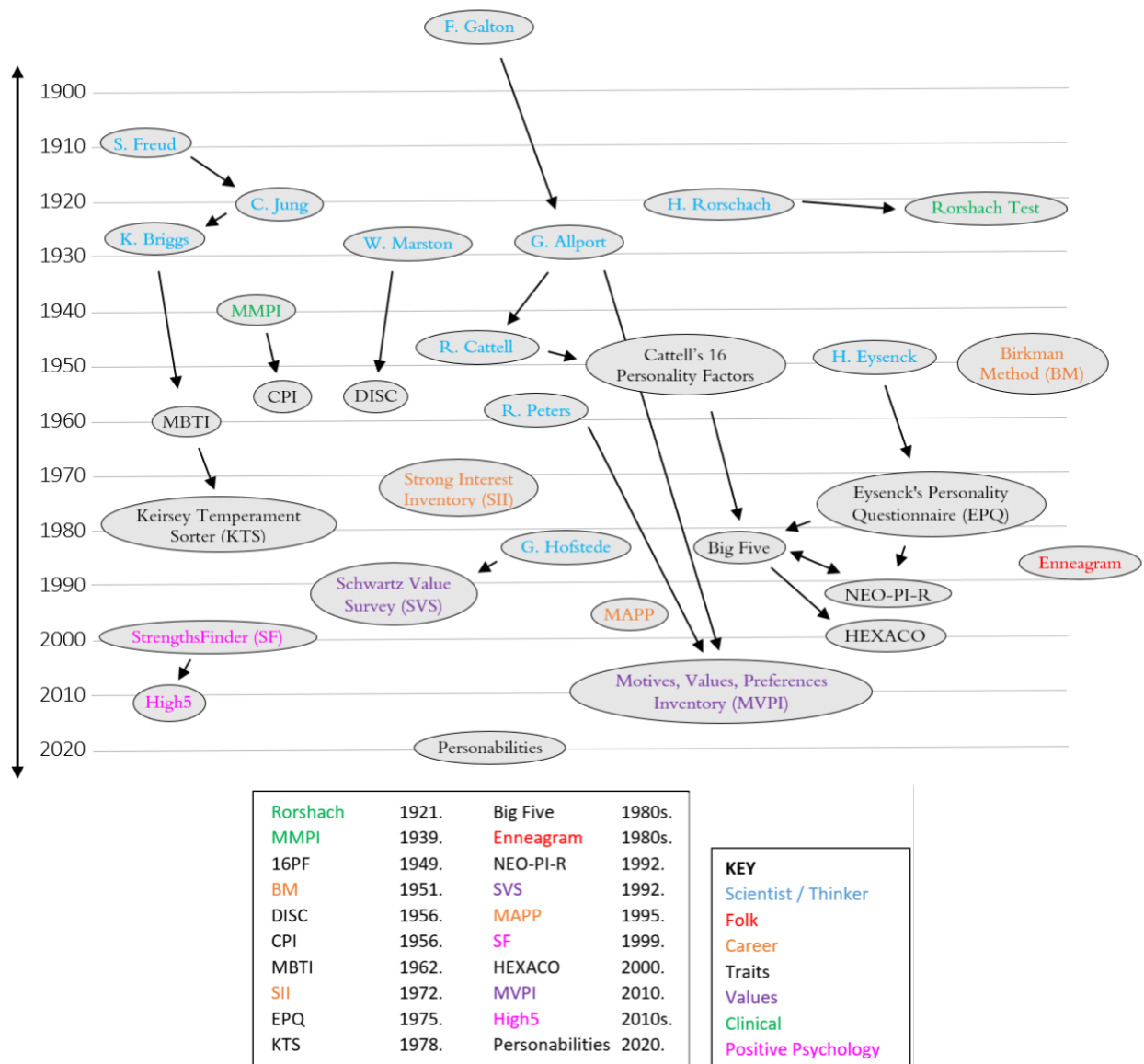Career
Traits
Values
Clinical
Positive Psychology

Figure 1. An attention space of the major personality models.

2. Dominant Design

Psychometric testing is split into two major categories: subjective and objective information collection.[15] Subjective information includes what participants think about, the emotions they experiences, and their worries/preoccupations. Some subjective assessment methods include:

1. The interview: aims to collect the participant's story by asking open-ended questions. The content (answers, habits, body language, etc.) is then analyzed.
2. Rating scales: people who know the participant rate him or her on several traits (e.g., a teacher rating a student from 1 to 5 on leadership capabilities).
3. Self-report tests: the participant completes a questionnaire, ranging from true vs false to more abstract forms such as an inkblot. Conclusions are drawn based off of answer patterns.
4. Projective techniques: aims to allow the participant freedom in expressing emotion and thought. The participant must interpret ambigious stimuli, such as pictures, text, art, etc.
5. Cognitive assessment: the participant records his or her thoughts and hypotheses at different points throughout the day, especially in regard to events or activities.

Objective information includes observable behavior, and some assessment methods include:

1. Behavioral assessment: conclusions are drawn about personality by assessing a person's observable behavior.
2. Bodily assessment: aims to reveal a person's feelings, motivations, and concerns by monitoring his or her physiology (e.g., heart rate, sweat, etc.) under certain conditions.
3. Personal facts: personality is assessed based on biographical data such as a person's history, text they've written, present activities, etc.

All of these subjective and objective psychometric techniques are utilized in personality psychology. However, the dominant technique, especially in the business world, is the use of self-report tests. These tests are also referred to as personality inventories, and one of the oldest and most popular examples is the Minnesota Multiphasic Personality Inventory (MMPI), a several-hundred-item questionnaire developed in 1939.
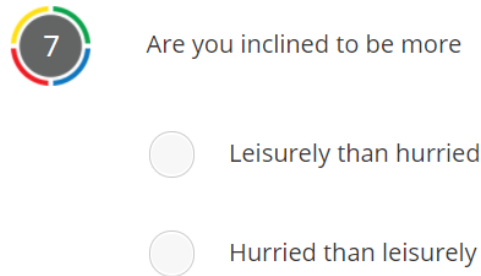
As a popular representative of the psychometric dominant design, I will examine the Keirsey Temperament Sorter. This assessment is used by more than 10,000 individuals daily, about 75% of Fortune 500 companies, and all branches of the US military.[16] Keirsey's business model operates primarily by holding workshops at corporations where employees can take the assessment

---

[15] I. G. Sarason, "Personality assessment", Encyclopedia Britannica, 1 Oct. 2019, https://www.britannica.com/science/personality-assessment. Accessed 5/22/2021.
[16] https://keirsey.com/. Accessed 5/23/2021.

and then learn about each other in order to work better as a team. The assessment consists of 80 questions in the following format:

Figure 2: a question from the Keirsey Temperament Sorter.

Keirsey then informs the participant of their temperament (Artisan, Guardian, Idealist, or Rational) and provides them with a report. The report is a lengthy PDF that includes information about the participant's personality and how it plays out in different parts of life, such as work. The general framework found in Keirsey is also true of MBTI, Enneagram, DISC, and emerging assessments such as Personabilites.

3. Technology Evolution

Interest in personality has grown exponentially in the computer field. Only about 15 papers with "personality" in their title were published in IEEE Xplore and ACM Digital Library (the two most important repositories of computing-oriented literature) in 2000. Since 2010, that number has grown to over 100 publications per year.[17] Interest in "personality computing" was fueled by three phenomena in the technological landscape:

1. The increasing amount of personal information available on social networking platforms.
2. The possibility of collecting everyday spontaneous, fine-grained behavioral evidence through mobile technologies, especially smart phones.
3. The attempt to endow machines with social and affective intelligence.

The field seeks to solve three problems: Automatic Personality Recognition (APR), Automatic Personality Perception (APP), and Automatic Personality Synthesis (APS). These three areas of research corelate to the Brunswik Lens, a model to explain how living beings gather information in the environment. The lens specifically describes the externalization and attribution of socially relevant characteristics during human-human and human-machine interactions. The intersection of personality computing and the Brunswik Lens can be seen below:

---

[17] A. Vinciarelli & G. Mohammadi, "A Survey of Personality Computing." *IEEE Transactions on Affective Computing* 5, no. 3 (2014): 273-91.
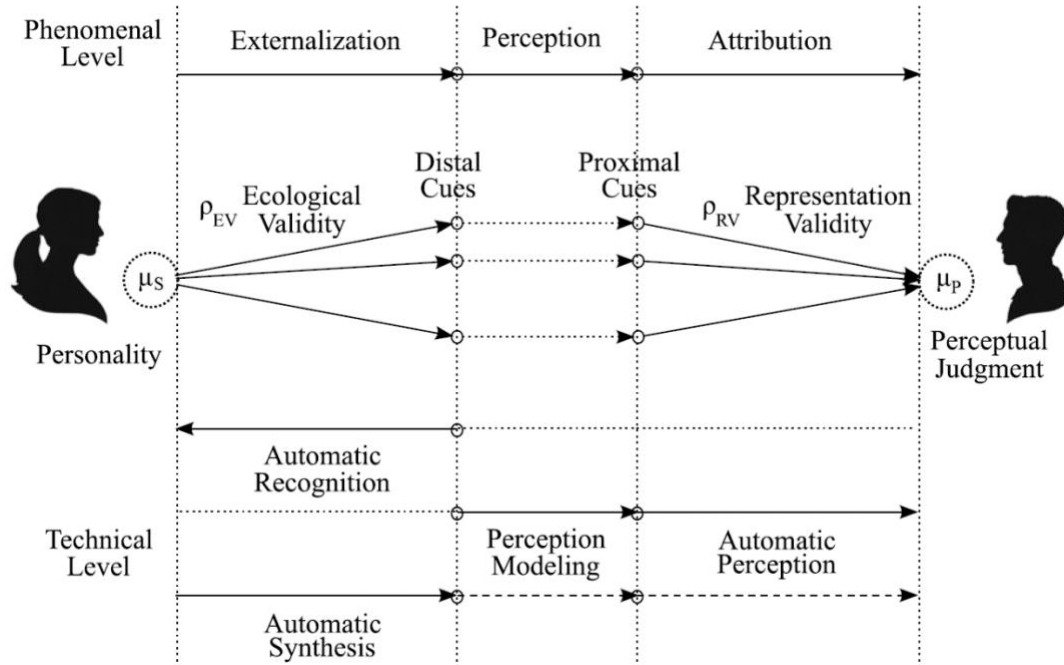
Figure 3: the relationship between the Brunswik Lens and the three main problems addressed in Personality Computing.[18]

There are a few key definitions. A *distal* cue includes any form of observable behavior perceived by others (human or machine). A *proximal* cue is what an observer actually perceives. The difference can be illustrated by speech: the acoustic waves emitted by a speaker are the *distal* cue and the perceived loudness is the *proximal* cue. *Ecological validity* describes any measure of the covariation between personality traits and distal cues. *Representation validity* describes the measures of the covariation between proximal cues and attributed personality traits. The three problems of personality computing can therefore be defined as:

1. <u>APR</u>: the task of inferring self-assessed personalities from machine detectable distal cues, targeting the externalization process.
2. <u>APP</u>: the task of inferring the personality observers' attribute to a given individual from proximal cues, targeting the attribution process.
3. <u>APS</u>: the task of automatically generating distal cues aimed at eliciting the attribution of desired (pre-defined) personality traits. The cues are generated in the externalization process, and the human assigns personality traits to the machine in the attribution process.

Extensive research has been done and continues to be done on APR, APP, and APS. However, since my research seeks to generate personality trait measurements via AI, I will focus on APR.

---

[18] Ibid, 276.

APR approaches primarily utilize trait-based personality models, in particular, the Big Five. The core distal cues that APR research has considered include written texts, non-verbal behavior, data collected via mobile or wearable devices, and online games.

4. Literature Review

Over the past fifteen years, computer programmers have attempted to predict personality using AI. The most prominent AI tool being used is natural language processing (NLP). The first step of NLP is to normalize a text, which normally includes three tasks:[19]

1. Tokenizing (segmenting) words
2. Normalizing word formats
3. Segmenting sentences

Through NLP, computers can recognize the syntax and semantics of language and draw connections between text.

In "A Systematic Literature Review of Personality Trait Classification from Textual Content," several scholars present over thirty attempts that were made at AI-powered personality prediction.[20] Many of the attempts performed NLP on large databases of text that were scraped from online social platforms such as Facebook, Reddit, and Twitter. The research focused on machine learning and deep learning approaches. These approaches and the associated algorithmic methods can be seen below:
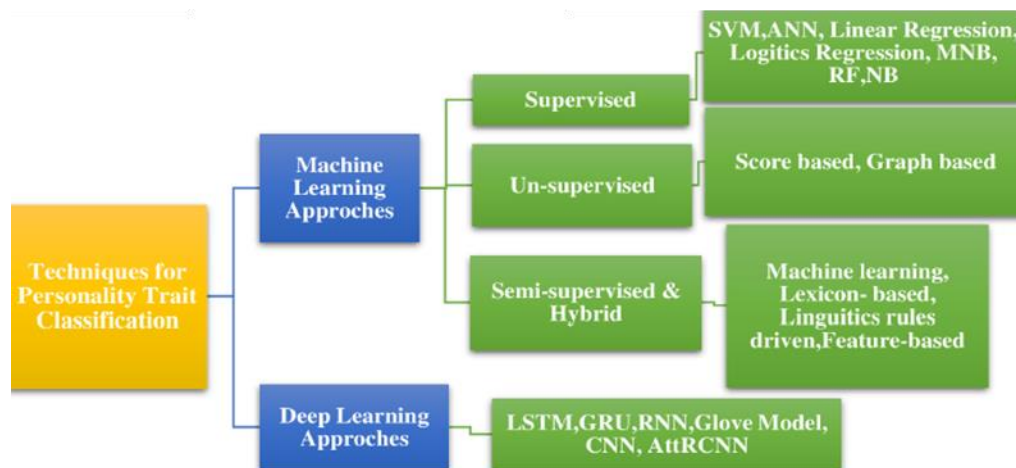

Figure X.

---

[19] D. Jurafsky & J. H. Martin*, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed. Pg. 14. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf. Accessed 5/14/2021.

[20] H. Ahmad, et al., "A Systematic Literature Review of Personality Trait Classification from Textual Content" *Open Computer Science*, vol. 10, no. 1, 2020, pp. 175-193. https://doi.org/10.1515/comp-2020-0188. Accessed 4/5/2021.

Most of the attempts to classify personality had over 80% accuracy rates which shows how much progress has been made in personality computing. Deep Learning in particular had impressive results with NLP on social media platforms.

Another significant piece of literature is a 2019 Sony Patent which seeks to use AI to predict MBTI types. The abstract states:

> This system may use various types of artificial intelligence to create arbitrarily granular detail of psychological typing so as to be able to apply it broadly. This modeling may be continuously updated and refined as the system learns about the degree of success of each of its recommendations.[21]

The patent further entails that Sony can monitor user activity and then market to the user based off of their calculated personality.

Research Questions

Unmet Needs

There are two major weaknesses that cripple the personality assessment industry. First, self-report questionnaires are an insufficient tool for producing consistently accurate personality trait measurements. In his thesis, James Nauert points out that, while self-reported data is helpful in capturing a test-taker's internal psychological process, "it can be difficult for individuals to know themselves well enough to provide an accurate image."[22] User responses often contain variability. Three things contribute to this variation:[23]

1. Random error. Participant's may experience changes in their mental state, such as attention levels and mental efficiency. This can result in a participant giving different responses to an item repeated at separate places in a questionnaire.
2. Transient error. Between two tests, events may impact the participant's mood, feelings, or mental processing ability.[24] Participants may then change a response in the second assessment.

---

[21] A. Galuten, (2019). *Identifying and Targeting Personality Types and Behaviors* (U.S. Patent No. 16/124,735). U.S. Patent and Trademark Office. https://bit.ly/3hOP4VG. Accessed 5/23/2021.

[22] J. Nauert, "Stable Inconsistency: A Study of Response Inconsistency Over Time," The University of Tennessee at Chattanooga (2017). Pg. 4. https://scholar.utc.edu/cgi/viewcontent.cgi?article=1663&context=theses. Accessed 5/14/2021.

[23] Ibid, 6. These three items originate with Schmidt, Le, and Ilies (2003).

[24] Nuart connects this to what Fleeson (2001) called state changes. "He argued that while personality can vary, responses on personality tests still provided a useful measurement. The scores we would typically associate with these personality traits were not static; rather they reflected a mean of behaviors in recent memory. These changes in behavior can be the result of internal processes, but are most visible when they are the result of an outside event" (Pg. 5).

3. Specific factor error. Participants may have unique reactions to the assessment itself. Factors like the wording of questions can influence a participant's response.

As a consequence of these errors, self-assessment is an insufficient means of providing accurate personality measurement.

The second weakness is that many personality assessments project limitations onto people. This is especially true of type models like MBTI. Many personality assessments will provide recommendations for a person based of their type. Here is an illustration of this.
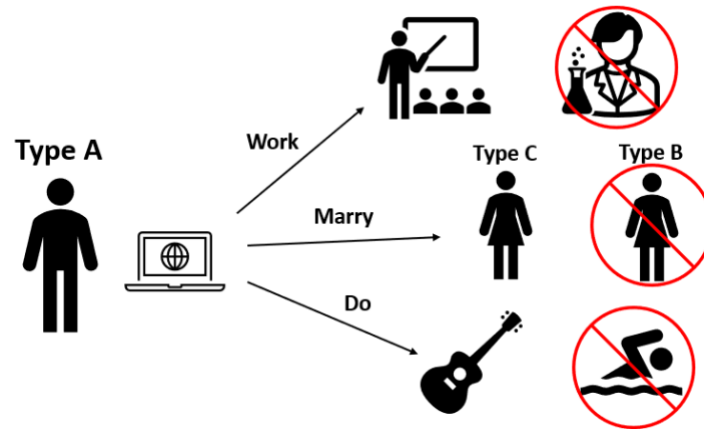

Figure X. Sample suggestions given based off of personality type.

By acting as the ultimate guide, personality corporations can influence participants to embrace a narrow view of their potential. This inhibits personal growth.

My capstone would seek to counter these two weaknesses. I would aim to solve the unreliability of self-report questionnaires by using AI to assess a technology user's personality traits more accurately. The data that the AI would collect over an extended period of time could then be averaged to give a holistic measurement of personality traits. I would then counter the second weakness of trajectory limitation by allowing users to plug in their own goals for the future. Materials would be selected to help them achieve those goals by taking the goal itself into account as well as their trait distribution.

Methodology

TBD

Testing

TBD

Deliverables

I am presently most impressed by the trait-based personality models since they have such high credibility in psychology research and are measurable. Although the Big Five is likely the most popular trait-based model, I am inclined to use the HEXACO model. Researchers Kibeom

Lee and Michael Ashton performed extensive factor analysis research in multiple languages and concluded that there is a sixth personality trait that transcends culture: the humility/honesty factor.[25] Thus, the traits of HEXACO are honesty/humility, emotionality (neuroticism in the Big Five), eXtraversion, agreeableness, conscientiousness, and openness. This model provides a holistic view of human personality, considering a person's darker side that may be indicated from a low score for H.[26]

I would therefore seek to predict HEXACO measurements using AI. I will discuss the details of what that could look like below.

Proof of Concept

Personality computation is extremely complex and will be difficult to perfect within the span of two semesters. I will thus define good, better, and best outcomes.

A good outcome will be to produce accurate APR for one trait within the HEXACO model, trait Z. I would need to define distal cues that a computer/smartphone user emits while using the technology. I would then need to quantify the covariation between those cues and trait Z and develop an AI program that is capable of recognizing and measuring those cues in order to update the value of trait Z.

A better outcome will be to produce accurate APR for one or more traits through the medium of a web or mobile application. The app would run AI in the background to monitor distal cues provided by the user, write down observations in a database, and update trait values based on the data. The user could view their trait measurements in the application along with information about those traits and what types of cues contributed to their score.

The best outcome would be an app that has accurate APR for all of the HEXACO traits and built-in personality information and growth resources. A user would be able to set goals and the app would provide them with recommendations for how to achieve those goals. In the process certain personality traits would increase or decrease in value. The model would look something like this:

---

[25] K. Lee & M. Ashton, *The H Factor of Personality: Why Some People Are Manipulative, Self-Entitled, Materialistic, and Exploitive-- and Why It Matters for Everyone*. Wilfrid Laurier University Press, 2012.

[26] K. Lee & M. Ashton, "The Dark Triad, the Big Five, and the HEXACO Model." *Personality and Individual Differences*, vol. 67, 2014, pp. 2–5.
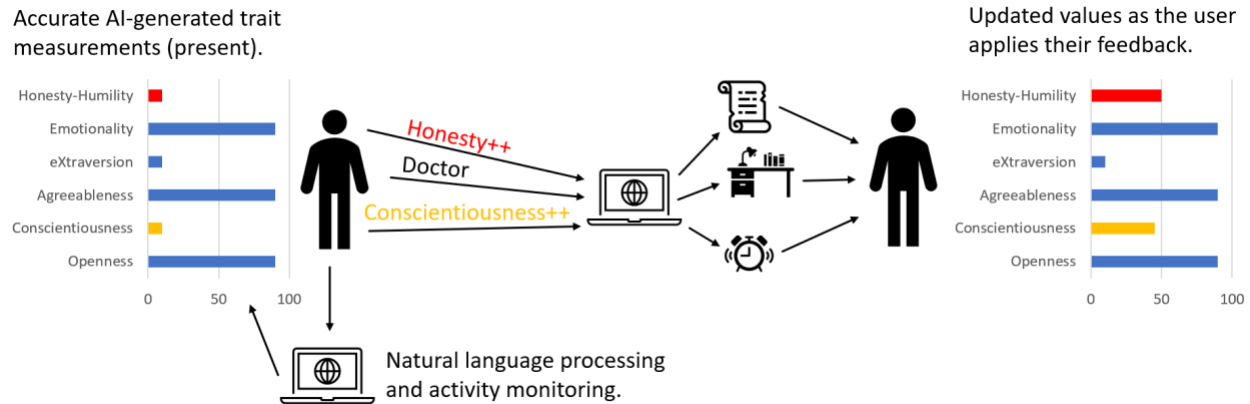
Figure X. An AI-powered personality assessment and growth tool.

Resources

Schedule

I would approach my capstone in four phases over the course of my senior year. The first two phases of research and prototyping would encompass the first semester, and prototyping would extend into the second. After finding an optimal prototyping I would move in the third phase of user testing and the fourth phase of revision and finalization. These phases are explored below:

1. Define the Metrics. Research the most accurate/useful personality traits to measure and which models are successfully doing so. Research the most effective deep-learning-based personality assessment methodology. Define how I will use the chosen deep-learning techniques to measure the chosen personality traits.
2. Create Prototypes. Write an AI program that can update a single personality trait measurement based off of distal cues. The initial distal cue would likely be text. Once I achieve some results with the first prototype, I would add more cues and possibly branch out to other traits.
3. User Testing. Develop a GUI that runs the AI software prototype. Recruit volunteers to take a self-report personality assessment and use the GUI. Monitor the personality trait measurement provided by the AI software and compare it to the measurement from the self-report.
4. Revisions and Finalization. Update the GUI and software based off of user feedback and measurement accuracy. Develop a final product that I can present and potentially release for use.

Cost

The cost is TBD. I foresee some expenses in acquiring certain AI tools and libraries but am still unaware of what those would be and their costs.

Partners

I would be interested in partnering with both researchers of AI and of psychometrics. In the Sattler community this could mean drawing on notes from David Anderson's research and from the counselling professors. Since the priority in my research is to test an AI program's ability to assess personality, it will be helpful for me to find a mentor in the field of AI.

Conclusion

I seek to use artificial intelligence to achieve automatic personality recognition with the HEXACO personality index. Personality is significant for every human, and creating an accurate online metric for personality traits could serve as a useful tool. I will focus my capstone on research and seek to accurately measure a single trait. However, if I have success early on with my good outcome, I will seek to develop a robust application that provides accurate personality measurement and tools for personal growth.