Applications of Artificial Intelligence to Information Privacy

James Frazier

Professor Tom Hemluth, ***Principal Advisor***

Professor Sally Cockburn, ***Secondary Advisor***

Professor Courtney Gibbons, ***Secondary Advisor***

Liberal Arts College
Hamilton College, New York
April 11, 2023

A proposal submitted to Hamilton College for Senior Fellowship

James Frazier

April 11, 2023

Key Words

Algorithms, Artificial Intelligence (AI), Computer Science (CS), Information Privacy, Security

Summary of the Proposal

Through Hamilton College's liberal arts education and its benefit of allowing a breadth of study, I have been able to experience opportunities and explore a myriad of topics that which would have not been possible through a traditional college experience. From various courses and research opportunities I have been a part of at Hamilton, my interests in artificial intelligence and information privacy have emerged. Artificial intelligence is the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages. Information privacy is the relationship between the collection and dissemination of data, technology, the public expectation of privacy, contextual information norms, and the legal and political issues surrounding them (Michael, 2013).

I aim, through this fellowship, to research and investigate the fields of artificial intelligence, information privacy, and ultimately the intersection of these fields with their modern-day advancements while continuing my on-campus commitments by utilizing the resources available to me at Hamilton College. Such an opportunity would be an appropriate conclusion to my undergraduate studies as the topic of my proposal aims to further my expertise and skills in fields which I hope to pursue a career in. Not only would this opportunity yield a more diverse and unique study for my senior year, it could also lead to new insights into such fields and/or the development of algorithms which utilize artificial intelligence to develop specialized cryptographic systems for increased security: a topic which has not been thoroughly explored compared to other uses of artificial intelligence.

1 Appropriateness of Fellowship

1.1 Contribution to the Liberal Arts Education and Career Goals

As a computer science major with a strong interest in artificial intelligence and information privacy – and by proxy cybersecurity – who hopes to pursue a career in these fields, I am interested in the potential of artificial intelligence to improve cryptographic (crypto) systems and enhance security/privacy features. A Senior Fellowship year would be a valuable addition to my liberal arts education and preparation for graduate school, allowing me to develop my research skills and deepen my knowledge in these fields, something not feasible with the current course selections at Hamilton. It would also provide me with an opportunity to work closely with faculty advisors, collaborate with other computer science and mathematics researchers, and potentially present my findings at a conference or publish them in an academic journal. Furthermore, it would give me the opportunity to gain hands-on experience in research, data analysis, and project management, which would be beneficial in both my future academic and professional pursuits.

1.2 Accomplishment of Breadth of Study

During my first three years at Hamilton, I took a variety of courses across multiple disciplines which can be found in Table 1.

Courses Taken at Hamilton College				
STEM			*Humanities*	
Computer Science	CS101, CS102, CS200, CS220	CS230, CS240, CS375,	History	HIST100W
Mathematics	MATH113, MATH116,	MATH216, MATH224W	German	GERMN110, GERMN120
Economics	ECON100,	ECON166	Literature	LIT208W
Physics	PHYS190,	PHYS245	Art	ART104, ART160
-	-	-	Philosophy	PHIL240
-	-	-	Japanese	JAPN110

Table 1: Courses I have taken at Hamilton College which are categorized by field classification and department.

By doing so, I have gained a broad understanding of the liberal arts and developed an appreciation for the interconnectivity of different fields. In particular, courses such as Symbolic Logic and Reasoning (**PHIL240**) helped me understand the applications of logic go beyond mathematics and the sciences and into other areas such as writing. Figure Drawing (**ART160**) illustrated how fields such as mathematics can be applied to the humanities to produce works while Murder, Civil War, and Opera (**HIST100W**) fostered an interest in the humanities overall.

I plan to supplement my interdisciplinary Senior Fellowship project on Applications of Artificial Intelligence To Information Privacy with coursework in related fields such as genetic programming with **CS411: Genetic Programming** and combinatorics with **MATH234:**

Counting and Codes while also studying various fields closely connected such as machine learning, data science, statistics, and mathematics. I also plan to investigate the social, structural, and institutional hierarchies that have been impacted historically by these fields. This will allow me to gain a deeper understanding of the broader applications and impact of artificial intelligence and privacy beyond my specific project, further contributing to my breadth of study. Overall, my combination of a broad selection of courses in my first three years and an interdisciplinary Senior Fellowship project will allow me to achieve the breadth of study that is a primary goal of a liberal arts education.

2 Outline of the Senior Year under Senior Fellowship

I aim to investigate the implementations and applications of artificial intelligence currently used in information privacy. The aim of this project is to develop an algorithm or a series of algorithms which utilize artificial intelligence to protect user information and privacy. Unless there are new findings which, at the discretion of the principal advisor Prof. Tom Helmuth, are deemed as necessary to include in my study, this section is anticipated to not change.

2.1 Previous Experience

During my undergraduate studies at Hamilton College, I have gained a strong foundation in computer science and mathematics, which will be essential in pursuing my proposed Senior Fellowship project. In particular, I have completed a course on artificial intelligence with Prof. Tom Helmuth, where I learned about various artificial intelligence techniques, such as neural networks, genetic algorithms, and decision trees. This course gave me a solid understanding of the theoretical underpinnings of artificial intelligence and its potential applications. Additionally, I have conducted summer-research into Human-Driven Genetic Programming (HDGP) with Prof. Tom Helmuth and other Hamilton students, where we explored the idea of incorporating human feedback into the genetic programming process; I have continued this research through this semester via the independent study course **CS298**. This research experience helped me develop my critical thinking and problem-solving skills in regards to artificial intelligence. Furthermore, I have attended numerous PUSH* Discourse meetings† with Prof. Tom Helmuth and other researchers, where we discussed topics related to artificial intelligence and other fields and concepts. These meetings helped me broaden my perspective on artificial intelligence and deepen my understanding of more complex concepts in the field. Finally, I have completed coursework in mathematics up to linear algebra and graph theory, which will be essential in understanding the mathematical concepts underlying cryptography, cybersecurity, and artificial intelligence.

2.2 Research Questions and Methodology

This Senior Fellowship aims to investigate the following questions:

1. What are the benchmarks of information privacy, and by proxy cybersecurity and cryptography, in regards to implementation?

*A stack-based programming language specifically designed for artificial intelligence to write code in.
†Virtual lab meetings

2. What are the applications of artificial intelligence in these fields? Furthermore, how did these implementations impact the fields?

3. Can artificial intelligence with its more recent advancements such as Lexicase Selection (Helmuth et al., 2016), stack-based GP (Stoffel and Spector, 1996), or Operator Equalization (Silva et al., 2011) be used in information privacy to get similar or even better results than current programs and/or systems in use today?

4. Are certain types of artificial intelligence (neural networks, genetic programming systems, decision trees, etc.) better suited for exploiting, discovering, and/or protecting user privacy? If so, which ones?

Research questions 1 and 2 are to build up the necessary skills and knowledge needed to investigate questions 3 and 4. Questions 3 and 4 will be the primary research questions I seek to answer during this fellowship.

In order to do so, I plan to spend the first semester of my fellowship studying artificial intelligence, cybersecurity, and information privacy separately to gain a deeper understanding of the fields on their own. Then, towards the end of the first semester and most certainly in the second semester, I plan to find possible applications of artificial intelligence in information privacy and implement these system to evaluate their effectiveness.

2.3 Related Work

Research into the applications of artificial intelligence to information privacy is scarce. Blackledge and Mosola (2020) provide a comprehensive analysis of the applications of artificial intelligence to cryptography. In their report, artificial intelligence has been successfully used in cryptanalysis[‡] and encryption. However, the applications considered in this report only include Machine Learning (ML) and Evolutionary Computing (EC) with the use of neural networks: genetic programming and other forms of artificial intelligence have not been considered because, while previous work such as the Lamar algorithm (Lamenca-Martinez et al., 2006) has found possible applications of genetic programming for example, they are not competitive with current advancements due to issues such as accuracy, scale, bloat, etc. which recent advancements such as Lexicase selection (Helmuth et al., 2016), Hyper-Heuristics (Kieffer et al., 2020), and Operator Equalization (Silva et al., 2011) have made progress to solve. In other words, the applications of artificial intelligence for protecting privacy has been discussed to minimal degree.

However, the utilization of artificial intelligence to exploit user information on the web is a well-known subject matter with growing concern in recent years, especially in concerns to user consumption (Mazurek and Małagocka, 2019) and healthcare privacy (Murdoch, 2021). Artificial intelligence has been applied to these fields to assist businesses and hospitals with sorting and analyzing large amounts of data with fast turn-around rates, but these same systems have exposed greater amounts of user information compared to their non-AI method and system counterparts. Furthermore, these artificial intelligence systems have been used to purposefully and effectively expose user information more often than it has been used to protect such information (Shank and Gott, 2020).

[‡]The practice of analyzing existing ciphers and protocols for weaknesses which occur as identifiable patterns that can be cloned.

2.4 Short and Long-term Objectives

The following are short-term objectives that I aim to complete during the first semester of my senior year:

- Independently create AI systems, crypto-algorithms, and privacy protocols which pass common benchmarks in their respective fields.
- Gain deeper understanding and knowledge of artificial intelligence.
- Investigate the current concerns regarding privacy.
- Design and implement a series of crypto-algorithms and privacy protocols.

The following are long-term objectives that I aim to complete after successful completion of the fellowship year:

- Understand the strengths and weaknesses currently in the fields of artificial intelligence, information privacy, cryptography, and cybersecurity and utilize them effectively.
- Successfully apply artificial intelligence such as genetic programming, neural networks, etc. to the field of information privacy that passes benchmarks for both fields.
- Research and produce artificial intelligence systems which aim to protect user information and privacy.
- Draft an extensive report on the utilization of artificial intelligence to enhance privacy of individuals.

2.5 Collaborators and Resources

2.5.1 Collaborators

I plan to collaborate with my advisors Prof. Tom Helmuth, Prof. Sally Cockburn, and Prof. Courtney Gibbons who specialize in artificial intelligence, graph theory, and number theory with applications to cryptography and cybersecurity respectively. Additionally, I plan to collaborate with various experts and researchers who frequently attend the Push Discourse virtual lab meetings such as Prof. Lee Spector of Amherst College. These individuals will be an invaluable source of expertise, knowledge, and advice on current practices in the field of artificial intelligence and the best methods for implementation.

2.5.2 Resources

I plan to fully utilize the expansive literature in Burke Library which contains various useful resources such as the book *Everyday cryptography fundamental principles and applications* (Martin, 2017) which will be necessary to learn the specifics of my study and stay informed of any new developments which may occur. I also plan to use the Computer Science Department's computer system Decker along with the High-Performance Computer (HPC) given to the college to utilize for computation-intensive research interests, all of which I already have access to at Hamilton College. Furthermore, open-source software[§] such as OpenSSH (ope) and implementations of information privacy and cryptography algorithms such as the *cryptopp* repository on GitHub[¶] (weidai11) will be utilized to assist in implementation and utilization of

[§]Software for which the original source code is made freely available and may be redistributed and modified.
[¶]An online platform that helps developers to collaborate on software development projects.

relevant software during this study.

2.6 Timeline

Table 2 is a comprehensive look into my projected study for the academic year of 2023-24 if approved for this fellowship. Below is an explanation of each of the subjects found in Table 2:

- **GP Systems**: Genetic Programming is a young field of artificial intelligence that implements and studies systems which mimic the process of natural selection in order to solve complex problems. In essence, it involves creating a population of computer programs that are randomly generated, and then using a process similar to natural selection to evolve and improve those programs over time. It has been used in areas such as evolutionary design, data mining, game design and optimization, image processing, and industrial control systems. This study will be an in depth look into the topic through Prof. Tom Helmuth's **CS411: Genetic Programming** seminar (more on this in Section 2.2.2).

- **Fundamentals of Crypto and Cybersecurity**: These studies will deal with protecting sensitive information and data stored and transmitted over the internet from unauthorized access, theft, or damage. It involves using complex mathematical algorithms to encrypt data in a way that only authorized individuals can decrypt and access the information. The field is concerned with developing and implementing measures to ensure the confidentiality, integrity, and availability of information and data in digital systems and networks.

- **Machine Learning**: Machine learning is a type of artificial intelligence that allows computers to learn from data and improve their performance over time. The idea is to give the computer large numbers of examples and let it figure out patterns or rules on its own, rather than having a programmer explicitly tell it what to do. I will study implementations and strategies related to such.

- **Statistics for Research**: Statistics play an important role in research by providing methods to analyze data. Here I will study the methods and algorithms that should be used in order to produce reliable and reasonable work and results.

- **Combinatorics**: Combinatorics is about figuring out how many different ways you can arrange or choose things. In other words, it deals with things like permutations, combinations, and probabilities. These concepts are used in many areas of science, engineering, and computer science to solve problems related to counting and arrangement. While I plan to study this subject during the Fall Semester independently, I also plan to take **MATH 234: Counting and Codes** in the Math department in the Spring semester to ensure a solid understanding of the topic has been achieved.

- **Fundamentals of Number Theory**: Number theory is the branch of mathematics that studies whole numbers, or integers. It is concerned with understanding the properties of numbers and how they behave when operations like addition, subtraction, multiplication, and division are performed.

- **Applications of AI to Info Privacy**: This subject will focus on what current applications of AI already exist for information privacy, how they fare, and how they have impacted the fields of information privacy, cybersecurity, and/or cryptography. Extensive literature

review into papers covering the topic will be done. I will also frequently consult my advisors on current developments (Prof. Tom Helmuth and Prof. Courtney Gibbons).

- **Implementation of AI for Use in Info Privacy**: This task will focus on implementing AI systems for use in information privacy based on the research and insights gained throughout the fellowship.
- **Drafting of Thesis**: This subject will focus on the drafting of my thesis which is required for completion of the fellowship.
- **Oral Defense Preparation**: This subject will focus on the drafting and preparations to orally defend my thesis as is required by the fellowship. Note, as the time of my oral defense is subject to the scheduling done by my principal advisor Prof. Tom Helmuth, this is subject's timeline will most likely change.
- **Investigate SSIH Impacted by AI and Privacy**: This subject will involve investigating the impact artificial intelligence, information privacy, and their intersections have had on social, structural, and institutional hierarchies. Note that this study occurs in the Spring semester of my Senior Fellowship so that I will have a good understanding of what systems and operations are when looking into their impacts.

The following are deadlines which I aim to meet during my fellowship that are not explicitly covered in Table 2.

- **Fall Semester - Week 8**: Produce several small and some large implementations of artificial intelligence and information privacy systems to report to principal advisor Prof. Tom Helmuth.
- **Fall Semester - Week 14**: Produce example implementations of machine learning artificial intelligence.
- **Fall Semester - Week 15**: Show proficiency in Number Theory, Combinatorics, and statistical research practices to secondary advisors Prof. Sally Cockburn and Prof. Courtney Gibbons.
- **Spring Semester - Week 3**: Produce a report for principal advisor Prof. Tom Helmuth over the social, structural, and institutional hierarchies impacted by artificial intelligence and information privacy.
- **Spring Semester - Week 9**: Have an implementation of an information privacy system which utilizes artificial intelligence which will be presented to principal advisor Prof. Tom Helmuth.
- **Spring Semester - Week 12**: Have my thesis finalized and begin preparation of my oral defense until my presentation. Note that as these timings are dependent on my principle advisor's scheduling of my oral defense, these are approximate timings of such events and are subject to change.

Fall Semester															
Task/Study	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Week 9**	**Week 10**	**Week 11**	**Week 12**	**Week 13**	**Week 14**	**Week 15**
GP Systems	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Fundimentals of Crypto and Cybersecurity	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Machine Learning	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Statistics for Research									X	X	X	X	X		
Combinatorics				X	X	X	X								
Fundamentials of Number Theory				X	X	X	X								
Applications of AI to Info Privacy									X	X	X	X	X	X	X
Spring Semester															
Task/Study	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 7**	**Week 8**	**Week 9**	**Week 10**	**Week 11**	**Week 12**	**Week 13**	**Week 14**	**Week 15**
Applications of AI to Info Privacy	X	X	X	X	X	X	X	X	X	X	X				
Implementation of AI for Use in Crypto and Info Privacy	X	X	X	X	X	X	X	X	X	X	X				
Drafting of Thesis				X	X	X	X	X	X	X	X				
Oral Defenese Preperation				X	X	X	X	X	X	X	X	X	X	X	X
Investigage SSIH Impacted by AI and Privacy	X	X													

Table 2: Timeline of my projected study for the 2023-24 Academic Year under the Senior Fellowship

Legend

Green:	Primary priority for the week
Yellow:	Secondary priority for the week
Red:	Smallest priority

2.7 Other Activities

The following lists other activities I plan to continue throughout my senior year under the fellowship:

- Continue to attend weekly Push Discourse virtual lab meetings to discuss current applications, publications, reports, and more in concerns to artificial intelligence.
- Present research insights and information to principal advisor Prof. Tom Helmuth in regularly scheduled meetings for guidance.
- Continue my current employments on campus
  - Student Research Assistant for Prof. Sarah Morrison-Smith
  - Computer Science Tutor
- Continue to support and participate in Hamilton AppDev events and projects
- Continue research into Human-Driven Genetic Programming (HDGP) with Prof. Tom Helmuth

2.8 Impact

This fellowship would allow me to explore an area which would otherwise be challenging with the current courses offered at Hamilton. It would also give me the opportunity to make a contribution to the scientific world and others while equipping me with the tools and knowledge necessary to make strides in the fields of artificial intelligence and information privacy. These fields will continue to grow as evident by tools, events, and phenomenon such as ChatGPT and massive security breaches become evermore frequent.

Additionally, this research could have possible applications in developing personal encryption systems where individuals can generate and create their own cryptographic protocols with the assistance of artificial intelligence to enhance their own privacy measures, putting the power of privacy into an individual's own hands. This research could also provide new insights to the fields of artificial intelligence and information privacy which could lead to others discovering new ideas and systems to utilize to increase the effectiveness of currently in-place systems. In other words, this research has the opportunity to investigate the cross-section of these fields and lead to broader applications of such.

References

https://www.openssh.com/.

Blackledge, J. and Mosola, N.: Applications of artificial intelligence to cryptography, Transactions on Machine Learning and Artificial Intelligence, 8, 21–60, https://doi.org/10.14738/tmlai.83.8219, 2020.

Helmuth, T., McPhee, N. F., and Spector, L.: Lexicase selection for program synthesis: A diversity analysis, Genetic Programming Theory and Practice XIII, p. 151–167, https://doi.org/10.1007/978-3-319-34223-8_9, 2016.

Kieffer, E., Danoy, G., Brust, M. R., Bouvry, P., and Nagih, A.: Tackling large-scale and combinatorial bi-level problems with a genetic programming hyper-heuristic, IEEE Transactions on Evolutionary Computation, 24, 44–56, https://doi.org/10.1109/tevc.2019.2906581, 2020.

Lamenca-Martinez, C., Hernandez-Castro, J. C., Estevez-Tapiador, J. M., and Ribagorda, A.: Lamar: A new pseudorandom number generator evolved by means of genetic programming, Parallel Problem Solving from Nature - PPSN IX, p. 850–859, https://doi.org/10.1007/11844297_86, 2006.

Martin, K. M.: Everyday cryptography: Fundamental principles and applications, Oxford university press, 2017.

Mazurek, G. and Małagocka, K.: Perception of privacy and data protection in the context of the development of artificial intelligence, Journal of Management Analytics, 6, 344–364, 2019.

Michael, M.: Uberveillance and the Social Implications of Microchip Implants: Emerging Technologies: Emerging Technologies, 2013.

Murdoch, B.: Privacy and artificial intelligence: challenges for protecting health information in a new era, BMC Medical Ethics, 22, 1–5, 2021.

Shank, D. B. and Gott, A.: Exposed by AIs! People personally witness artificial intelligence exposing personal information and exposing people to undesirable content, International Journal of Human–Computer Interaction, 36, 1636–1645, 2020.

Silva, S., Dignum, S., and Vanneschi, L.: Operator equalisation for Bloat Free Genetic programming and a survey of Bloat Control Methods, Genetic Programming and Evolvable Machines, 13, 197–238, https://doi.org/10.1007/s10710-011-9150-5, 2011.

Stoffel, K. and Spector, L.: High-performance, parallel, Stack-based genetic programming, Genetic Programming 1996, https://doi.org/10.7551/mitpress/3242.003.0030, 1996.

weidai11: Weidai11/cryptopp: Free C++ class library of cryptographic schemes, https://github.com/weidai11/cryptopp.