



DISSERTATION TOPICS

MSc/MPhil in Information Technology

This document lists the topics suggested by supervisors in the School of Information Technology. For full guidelines on selecting a supervisor, please see the Dissertation Guidelines.

DISSERTATION RESOURCES

- [Full listing of School of IT staff](#) (note that Adjunct and Honourary staff do not typically supervise MIT students)
- [Examination Instructions](#)
- [Faculty of Science Research Ethics](#) (for Human Subjects Research)
- [CS Dissertation Archive](#)

TABLE OF CONTENTS

Dissertation Resources	1
Krupa Prag (Computer Science)	3
Maureen Tanner (Information Systems)	3
Tendani Chimboza (Information Systems)	3
Maria Keet (Computer Science)	4
Generation of semantics-enhanced text for adaptive e-learning	4
Interactive template design	4
Prospects of cross-language reuse of spellcheckers for African languages	5
Michelle Kuttel (Computer Science)	5
BiofilmBuilder: developing a tool to build models of bacterial biofilms	5
Glycan3DB: developing a comprehensive database of carbohydrate structure and conformation	5
Designing an effective interactive visualisation for a complex data set	6
Geoff Nitschke (Computer Science)	6
Artificial Neural Networks for soccer agents	6
James Gain (Computer Science)	6
Comparing Sketching Techniques	6



Hussein Suleman (Computer Science)	7
Recommending Digital Repositories (Hussein Suleman)	7
Electronic heritage resources in local languages (Hussein Suleman)	7
Rob Simmonds (Computer Science)	7
Managing Cluster Invocation for Remote Astronomy Visualisation	7
Extending Scripting for Astronomy Visualisation.....	7
Melissa Densmore (Computer Science).....	8
AI-enabled Hydroponic Farming	8
Multimedia Peer Support for Breastfeeding Mothers and Early Childhood Development	8
Alternative Network Infrastructures for Data-Constrained Internet Use.....	8
Co-Design for English Language Learning	8
Evaluation of SUS in Multilingual Settings	9
Co-Design and Evaluation of Local Services for Community Wireless.....	9
Tommie Meyer (Computer Science).....	9
Lightweight Ontology Editing and Reasoning	9
Josiah Chavula (Computer Science)	9
State of Local Internet (Josiah Chavula).....	9
Jan Buys (Computer Science).....	10
Neural Machine Translation for African Languages.....	10
Deep Learning-based Text Analysis for South African Languages	10
Incorporating Graph-based Representations in Neural Dialog Systems (Jan Buys).....	10
Patrick Marais (Computer Science)	11
Detecting missing trees in drone-based orchard survey images	11
Determining local row orientation from drone-based orchard survey images	11
Zola Mahlaza (Computer Science)	11
A controlled natural language for research questions in Computing	11
Francois Meyer (Computer Science).....	12



Aligning Word Boundaries for Multilingual Machine Translation	12
UCT Knowledge Coop	12
Wallace Chigona (Information Systems).....	13
Zainab Ruhwanya (Information Systems).....	13
Cybersecurity And Privacy Research Group (CSPR)	13

KRUPA PRAG (COMPUTER SCIENCE)

1. From [UCT Knowledge Co-op](#): # 810. Optimising Plant Selection for Urban Greening in Cape Town.
 - Using Computational Intelligent techniques and data analysis.
2. Optimisation of the efficiency of a DC-DC Buck-Boost Converter
 - Optimise the accuracy, stability and efficiency of a DC-DC Buck-Boost Converter using Reinforcement learning and Computational Intelligent techniques.
3. Optimising the Vehicle Routing Problem using Computational Intelligent Techniques, Reinforcement Learning and Quantum Computing.
 - Optimise the Vehicle Routing Problem concerning number of vehicles and total distance using Computational Intelligent Techniques, Reinforcement Learning and Quantum Computing.

MAUREEN TANNER (INFORMATION SYSTEMS)

1. How can software development teams leverage AI technologies to create solutions that address marginalized communities' specific needs and challenges, contributing to their social and economic advancement?
2. How can software development teams leverage AI technologies to create solutions that address the specific needs and challenges faced by marginalized communities in the Global South, contributing to their social and economic advancement?
3. How might the development and implementation of AI governance frameworks impact software development teams' structure, decision-making processes, and global collaboration efforts?

TENDANI CHIMBOZA (INFORMATION SYSTEMS)

1. Rural development
 - a. Digital rural community initiatives for employment creation
 - b. Mapping technology availability and use in rural FETs.



2. Cybersecurity
 - a. Cybersecurity awareness in rural areas: The case of Africa countries
 - b. Internal cybersecurity policy compliance for African organisations
 - c. Government commitment to cyber resilience in Africa
 - d. Do schools in rural areas have cybersecurity rules?
 - e. cybersecurity behaviour of often neglected society (senior citizens)
 - f. Cybersecurity framework for teachers in rural areas
3. Technology and children
 - a. Monitoring and regulating children's online behaviour through Parental Control Software (PCS)
 - b. Awareness of PCS in rural areas
 - c. Digital parenting behaviours in South Africa.
4. Cyberpsychology
 - a. Twitter (X) algorithm that affects mental wellbeing of Adults
 - b. Geographical resistance of digital monoculture
 - c. Naive recklessness on social media posting: Consequences on influencers
 - d. Monetization models of social media entrepreneurship.
5. Digital ethics
 - a. Ethical issues associated with collecting social media data
 - b. Algorithmic injustice that perpetuate marginalisation in social media
 - c. Lack of ethical ownership on applications that enable digital gaslighting and cancel culture

MARIA KEET (COMPUTER SCIENCE)

Generation of semantics-enhanced text for adaptive e-learning

Most textbooks are offered as hardcopies or plain pdf files where the sections, figures, and tables are hyperlinked at best. The notion of **adaptive e-learning** aims to offer more than just a softcopy, and instead seeks to incorporate other technologies into the softcopy of a textbook to provide a better learning experience. Such techniques include context-sensitive automated question generation and marking of the answers. Textbook-independent tooling is scarce, however; the only knowledge-driven showcase with such features ("Inquire biology") is proprietary, so we are building one in-house. You will investigate how best to automatically annotate text with a knowledge graph so that various types of context-sensitive questions can be generated and document navigation can be improved.

Interactive template design

Wikipedia is one of the most popular sites on the Web. Articles are written from scratch for each supported language, which is very resource-consuming. A new proposal, and Wikimedia Foundation Project, was launched in 2020 to address that: Abstract Wikipedia. It aims to use Wikidata as content source and Wikifunctions as algorithm store to generate articles automatically, in any natural language. You will investigate one of the prospective tasks, being automation surrounding the writing of templates with the template language that we defined recently (https://meta.wikimedia.org/wiki/Abstract_Wikipedia/Template_Language_for_Wikifunctions).

Conceptual model design and validation techniques

Basic design of ER and UML Class diagrams in computer science courses is largely based on paper or a simple drawing tool to cover the basics. That practice does not help students to design good models. Disparate theory and techniques have been proposed in the literature to streamline the design process of drawing the diagrams,



such as the test-driven model design, automated reasoning to detect inconsistencies, and a conceptual schema design procedure. Software support is lacking for that and suspected to be the reason that such aids are not used. For some, it is not clear to what extent such techniques improve modelling or the learning of modelling. The aim is to design and implement one or two of such advanced methods or methodologies, or at least component methods, to enhance conceptual model design and evaluate whether it indeed makes a difference, and if so, where and how.

Prospects of cross-language reuse of spellcheckers for African languages

Effective computational support for document preparation in most natural languages requires error prevention, detection, and correction. First successes for detection and correction have been obtained with data-driven spellcheckers for isiZulu and isiXhosa using a statistical language model for error detection and limited error correction. Among the remaining questions are how much data is needed to create a good spellchecker and how easy, or hard, it may be to transfer the current set-up to other Niger-Congo B languages, be they closely related siSwati or isiNdebele, somewhat related ciShona, or a language further afield or with different orthography, such as Setswana, or a dialect. Your task is to address either of the topics. Note: if a statistical approach is taken, no knowledge of the languages is required; if a rules-based approach is taken, it is highly recommended you know at least the basics (i.e., at least up to matric-level) of the selected language(s).

MICHELLE KUTTEL (COMPUTER SCIENCE)

BiofilmBuilder: developing a tool to build models of bacterial biofilms

We have an on-going project to develop scientific software tools to aid research into the structure and function of carbohydrates (molecules that play key roles in disease such as HIV and TB and are the focus of many modern vaccines). We have a publicly available console tool for building carbohydrate models called CarbBuilder [1].

This challenging project would use this tool to develop a BiofilmBuilder tool. Many microorganisms form biofilms: an external molecular matrix that keeps the bacterial cells in close proximity and protects the bacteria from attack. Biofilms are implicated in many bacterial infections (e.g. dental plaque is a biofilm). Polysaccharides are among the most abundant components of the biofilm, but they also contain proteins and other molecules - and the 3D structure of any biofilm at the molecular level has not yet been modelled. This project would leverage or extend our CarbBuilder software tool for producing three dimensional molecular models of carbohydrates to automatically build large models of biofilms.

1. CarbBuilder: Software for Building Molecular Models of Complex Oligo- and Polysaccharide Structures. M. M. Kuttel, Jonas Stahle, and Goran Widmalm, *J. Comput. Chem.*, 37(22): 2098-2105 (2016). DOI: 10.1002/jcc.24428.

Glycan3DB: developing a comprehensive database of carbohydrate structure and conformation

We have an on-going project to develop scientific software tools to aid research into the structure and function of carbohydrates (molecules that play key roles in disease such as HIV and TB, and are the focus of many modern vaccines). We have two publicly available web tools for investigating carbohydrate structure: Glycarbo [1] and Glycan3DB [2].



Glycan3DB is currently a searchable database of carbohydrate structure for the organism *Klebsiella pneumoniae*. We would like to extend this from a first prototype to a fully functional and usable database that will cover a wider range of bacteria and have additional functionality. Searches for complex structure fragments within the database should be supported. Uploading and curating data is a time-consuming process and Version 2 should allow for frequent actions to be automated. The possibility to scrape paper repositories for structure data should also be explored.

1. <https://glycarbo.cs.uct.ac.za>
2. <https://glycan3db.cs.uct.ac.za/>

Designing an effective interactive visualisation for a complex data set

Visualisation of complex data is increasingly important, both for enabling researchers to explore data effectively to identify trends, correlations etc., as well as for communication of complex information. The prevalence of complex charts during the Covid-19 pandemic is a case in point. Designing an effective interactive visualisation for a complex data set requires not only technical skill, but an understanding of human perception processes as well as the methods for effective design. However, many visualisations are ineffective or, worse, biased and untruthful. This project will **require students** to identify a **suitable complex dataset** (preferably one they are familiar with and interested in) and then go through a rigorous **iterative process of design** to produce a final system, which should then be **tested for efficacy**.

GEOFF NITSCHKE (COMPUTER SCIENCE)

Artificial Neural Networks for soccer agents

The topic falls into the field of Artificial Intelligence (AI). The topic is to implement an AI method that automatically teaches simulated soccer agents to kick and intercept a ball on a virtual soccer field. That is, maximizing the number of successful passes and intercepts using a learning approach. The soccer agents will be controlled with an Artificial Neural Network (ANN) and trained with the back-propagation algorithm. This training process is analogous to a coach providing feedback (good or bad) and many training examples from which players learn the desired skill (passing or intercepting). The ANN controllers and back-propagation algorithm will be implemented in the context of an existing Java based soccer simulator. The goal is for the soccer agents to begin with no skills (untrained), but after many simulated training examples, one soccer agent will be trained to intercept balls, and the another to score goals, with a high degree of accuracy.

JAMES GAIN (COMPUTER SCIENCE)

Comparing Sketching Techniques

Many computerised drawing and modelling applications require a user to draw smooth curves, either with a mouse or pen and tablet. There are several different approaches to doing this. For instance, one technique allows users to position points (with optional tangents at the points) and have a smooth curve fitted through the points. Another approach is to allow them to draw free-hand and then smooth the resulting curve afterwards. However, there has been little in the way of assessment of the strengths and weaknesses of different approaches. This study will consider the options available for drawing such curves, implement a number of them and then compare them experimentally by running one or more user studies.



HUSSEIN SULEMAN (COMPUTER SCIENCE)

Full list of topics/ideas by Hussein Suleman are at:

<http://www.husseinsspace.com/research/prospective.html>

Recommending Digital Repositories (Hussein Suleman)

Develop and evaluate a multi-criteria recommendation engine for selecting a digital repository technology. Selecting the most appropriate digital repository tool (e.g., DSpace vs. EPrints) is somewhat of a black art because of the large number of variables and the difficulty in making direct comparisons among tools. Nevertheless, many repository managers have made successful choices, while some have also made poor choices. Using this data (which will need to be collected), patterns can be learnt using machine learning to help repository managers make better choices in future.

Electronic heritage resources in local languages (Hussein Suleman)

The Bleek and Lloyd Collection of early South African heritage artefacts preserves the culture of an important group of South Africans. The online archive provides users with copies of notebooks containing stories in English, |xam and !kun. The direct descendants of these groups do not, however, speak any of these languages so the information is not useful to them without translation and reinterpretation. Simultaneously, the reinterpretation and translation may provide useful information to researchers studying the original texts. Two recently completed projects looked into using AI and crowdsourcing to create transcriptions of the text, while previous work from researchers included manual transcription. This project will look into how to present these multiple representations of the information so that users viewing the information can use the most appropriate representation and/or contribute/enhance the existing information. Examples of similar projects are Google Goggles and how Chrome does translation of Web pages. The work will make research contributions in the area of novel interfaces for the creation, display and enhancement of heritage information.

ROB SIMMONDS (COMPUTER SCIENCE)

Managing Cluster Invocation for Remote Astronomy Visualisation

This project involves considering how to use Kubernetes to deploy the backend of a remote visualisation system onto computer systems. It will include cases where the backend is using shared memory and where it is using a data-flow framework to run on multiple compute nodes. As a base case the CARTA remote visualization system that is developed in part at UCT can be used to explore a specific use case.

Extending Scripting for Astronomy Visualisation

CARTA is a remote visualisation system developed in part at UCT. It enables users to access large radio astronomy data cubes that are housed in cloud and HPC systems. It has recently had a scripting interface added that enables users to drive the visualisation system from applications such as Jupyter notebooks. This project involves expanding support for additional use cases, but where the



visualisation session is viewed by remote users and when the system is used to generate images for other applications.

MELISSA DENSMORE (COMPUTER SCIENCE)

AI-enabled Hydroponic Farming

Food security is a key development issue in South Africa, exacerbated by rising unemployment and climate change. We have worked with farmers in KwaZulu-Natal to develop a system using AI processing and sensors attached to an Arduino board to support hydroponic farming. This empowers farmers to grow food in spite of changing weather conditions or limited space. For this research we will work with urban farmers in Cape Town to think about how this tent might be used to meet the needs of the community, using a co-design approach to explore the use of AI tools in community-centred settings. This will entail at least two weeks of fieldwork in Ocean View and regular meetings with the our collaborators for the duration of your work.

Multimedia Peer Support for Breastfeeding Mothers and Early Childhood Development

For this research we will explore the ways in which data-constrained parents of children in the first thousand days (conception through 2 years) can produce, share and use multimedia content that will help them adopt best practices, as well as provide mental support for a stressful time. Students proposing to do this project will work with one of our existing populations (mothers of infants in the NICU, breastmilk donors, township residents) or can propose their own population and context for the research.

Alternative Network Infrastructures for Data-Constrained Internet Use

There is a growing population of people whose primary, and often only computing device is a mobile phone, for which internet connectivity can be unreliable and prohibitively expensive. At the same time, many Internet-based applications and service presume access to unlimited bandwidth and experience with computers. In this research you will seek to better understand this population of users (and their intermediaries) and the ways in which they use mobile phones, data, and the internet. questions around how to motivate participation and interactivity. Your work may consist entirely of measurements, providing evidence for the ways in which existing applications might constrain digital participation, or you may work with your population to co-design tools to help people manage bandwidth, make more effective use of the Internet or any other purpose you develop with the users.

Co-Design for English Language Learning

One of the most significant barriers to employment for residents of Cape Town is English language proficiency. Many organizations offer English classes to support job seekers. However, students find the often volunteer-driven classes to be insufficient. While many use mobile apps such as dictionaries or translators to support their learning others do not, either to availability of supporting devices or limited access to the internet. In prior studies, students have expressed interest in pictorial and conversational apps to support the development of proficiency in English. For this study, you will be expected to explore the space of apps and services for English language learning, and to work with students to co-design something to support their studies. This may take the form of app-based vocabulary learning, a WhatsApp-based conversation/mentorship program, or any other ideas you are able to co-design with English language learners. Fluency in isiXhosa or French is an advantage for this topic.



Evaluation of SUS in Multilingual Settings

The System Usability Scale is a widely used metric for assessing the usability of software. It is easy to use and shown to be effective. However, some research has shown that poor translation across languages can affect its effectiveness. In multi-lingual contexts this effect can be exacerbated since often participants are not completing the scale in their home language. For this project you will evaluate SUS in a multi-lingual setting and demonstrate whether there's a difference in the efficacy, precision and accuracy of SUS when the scale is conducted in home language vs in a common language, or in multi-lingual contexts.

Co-Design and Evaluation of Local Services for Community Wireless

We are working with the community of Ocean View (South Peninsula, Cape Town) to build a community wireless network to enable low-cost local communications services. For this research, you might help develop and test some aspect of our software (such as content synchronization or usable authentication) or co-design a service (such as music authoring, community radio, chat). This will entail at least two weeks of fieldwork in Ocean View, and regular weekly meetings with the Ocean View team for the duration of your work.

TOMMIE MEYER (COMPUTER SCIENCE)

Lightweight Ontology Editing and Reasoning

Protégé is a free and open-source ontology editor. Using Protégé, it is possible to employ logical reasoning to build high-quality ontologies and to query them, thereby completing the first step in building intelligent systems. Its plugin architecture allows for extending the basic capabilities built into the core Protégé system.

The goal of this project is to extend the capabilities underlying the core Protégé system. The exact nature of the extension is open for discussion, ranging from the provision of better user interfaces for the existing Protégé, to the integration of non-standard forms of reasoning into the Protégé environment. Another possibility is to dispense with Protégé altogether, and develop a custom-made ontology-editing tool, with access to logical reasoners through the OWL API.

JOSIAH CHAVULA (COMPUTER SCIENCE)

State of Local Internet (Josiah Chavula)

For Internet users, it is often useful to be able to compare Internet access options within specific/localized geographical regions. In this project, we collect the necessary sample datasets, and build a web-based visualization to allow easy exploration of various types of local Internet datasets to help communities better understand the state of connectivity in specific local jurisdictions (e.g. province, municipality, ward). This will extend on the work that we have collaborated with ISOC-ZA and AFRINIC to build the World Internet Data Explorer (WIDER).

This project will undertake user-centric network measurements with a particular city/town/district (eg Cape Town), with an interest to understand how users in low-resource settings (such as in community networks, townships, public wifi) experience and perceive individual services and applications (e.g., video/audio streaming, gaming, conferencing, etc.) over different terminals (e.g., mobile phones, tablets, and computers), and over different networks and locations. Since these services have different requirements on performance and



application-level metrics, their perceived performance need to be assessed separately. Among others, the project will curate, aggregate and visualize Internet access and performance data obtained within a city. The data collected in this exercise will contain geo-location information, allowing localized visualizations, giving communities the ability to view Internet metrics at granular geographic levels. Geographic visualizations could include availability and quality of service for different networks, such as mobile operators. Additional/complementary datasets will be obtained from M-Lab dataset through the Google BigQuery and Ripe Atlas among others.

JAN BUYS (COMPUTER SCIENCE)

Neural Machine Translation for African Languages

Deep Learning has led to substantial advances in automatic translation quality. However, only a limited number of neural machine translation (NMT) systems have been developed for African languages, and small data sizes and typological differences compared to Western languages pose unique challenges. The goal of this project is to pick an African language (or a few closely related ones) for which no openly available translation system exists, find training data and train a standard NMT system. The system should then be improved to optimize performance for the given language(s). Possible approaches that can be investigated include: Better data preprocessing methods (e.g. using Byte Pair Encoding), comparing different neural network architectures, applying data augmentation techniques (e.g. backtranslation or multilingual training), and optimizing performance for a specific domain (e.g. for medical text).

Basic knowledge of machine learning is recommended. Knowledge of an African language will be helpful but is not required. The project will involve using existing NMT toolkits, so implementing neural network models will not be required.

Deep Learning-based Text Analysis for South African Languages

Natural Language Processing (NLP) systems often require linguistic analysis of text. These tasks include lemmatization, parts-of-speech tagging, phrase chunking and named entity recognition. There has been limited development of systems that perform these tasks for South African languages. The goal of this project is to develop neural network-based models for all or a subset of these tasks for multiple South African languages. Existing code bases will be used to train and test models, but data preprocessing and tuning and comparing different models will be required. As many South African languages are closely related, multilingual training methods will also be explored. Basic knowledge of machine learning is recommended. Prior experience in NLP or linguistics is not required.

Incorporating Graph-based Representations in Neural Dialog Systems (Jan Buys)

Existing Neural Network-based dialog systems (chatbots) often generate responses that are inappropriate in the context of the dialog or that contradicts given information or what was previous said in the conversation. This project will explore the use of graph-based representations of the content of the dialog and of information related to the dialog in order to improve the faithfulness of the system outputs. Information related to the dialog will be represented as a Knowledge Graph, and candidate outputs generated by the neural network model will be parsed to semantic graphs (using Abstract Meaning Representation (AMR)) and verified for relevance and consistency. A representation of the chosen output will then be added to the Knowledge Graph. An optional



second part of this project will involve extending the neural network model to condition the response generation on the Knowledge Graph or graph representations of the dialog.

Some familiarity with neural networks is recommended. An existing neural dialog system and AMR parser will be used, but some programming will be required incorporate the systems (and in the second part to retrain the neural network).

PATRICK MARAIS (COMPUTER SCIENCE)

Detecting missing trees in drone-based orchard survey images

Precision agriculture is a relatively recent concept that seeks to leverage IT to bolster crop production, amongst other things. In this project we have access to drone image surveys of orchards, which are either regular RGB images or RGB images with additional data per pixel. The goal of the project is to detect regions where trees may be missing in orchard rows. This is a very challenging problem, since it requires detecting 'gaps' in rows, and the rows themselves are often not straight or regular, but follow land contours, and may not even be clear visually. One possible approach is to use heuristic-based classical image processing methods. It is strongly suggested that anyone tackling this project have experience with technical Python programming (perhaps from data science). Exposure to image processing methods and concepts would be very helpful and is more likely to lead to the development of a successful solution.

Determining local row orientation from drone-based orchard survey images

In this project we have access to drone image surveys of orchards, which are regular RGB (colour) images. The goal of the project is to provide an estimate of the local orientation of rows across the orchard. In this context, "local orientation" refers to the angle that the orchard row makes (0-180 degrees) within a small neighborhood around each tree. This requires determining a method to locally estimate the row orientation, which is not a trivial problem. This is useful information for high-level algorithms which attempt to automatically identify complete rows of trees from each image. It is strongly suggested that anyone tackling this project have experience with technical Python programming (perhaps from data science). Exposure to image processing methods and concepts would be very helpful and is more likely to lead to the development of a successful solution.

ZOLA MAHLAZA (COMPUTER SCIENCE)

A controlled natural language for research questions in Computing

Learning to write good research questions in Computing fields is a difficult task for students. Other fields, especially ones related to medicine, use frameworks such as PICO (Population, Intervention, Comparison, Outcome) to assist in such an endeavour, in addition to structuring systematic literature reviews. These frameworks are rarely employed in Computing and that is likely because such tools specify what ought to be included in questions and do not specify how to author them. There is a need to create a controlled natural language (CNL), and associated authoring tool, that can be used to create good research questions. In this project, you will design a CNL for a chosen field that lies within Computing and evaluate the extent to which it yields to good research questions.



FRANCOIS MEYER (COMPUTER SCIENCE)

Aligning Word Boundaries for Multilingual Machine Translation

Deep learning models have proven highly effective as machine translation systems but require large amounts of data. One method to improve translation for data-scarce languages is to develop multilingual models. These models are trained to translate between multiple input and output languages. This enables cross-lingual transfer, where similarities between languages allow the model to utilise its knowledge about one language to improve performance for another. The languages of South Africa provide an interesting case for experimenting with cross-lingual transfer. The Nguni languages (e.g. isiXhosa) and Sotho-Tswana languages (e.g. Setswana) exhibit distinct orthographic word boundary conventions—the level of meaning at which words are separated by spaces in the writing system. IsiXhosa words are often quite long, with many single isiXhosa words translating to multiple Setswana words. This mismatch can prevent cross-lingual transfer between isiXhosa and Setswana, despite underlying similarities that would otherwise facilitate transfer.

This project aims to overcome these challenges by aligning word boundaries in multilingual machine translation. One approach to aligning word boundaries is to pre-segment isiXhosa words into morphemes (meaningful subword units), making the written words more comparable in length and meaning to Setswana words. By aligning the orthographic word boundary conventions between isiXhosa and Setswana, we can potentially enhance cross-lingual transfer, improve translation accuracy, and better understand the trade-offs of multilingual modelling.

Basic knowledge of machine learning is recommended. Knowledge of the South African languages is not required, as the project will be fully data-driven. The project will involve using existing Python toolkits for machine translation and deep learning, so implementing neural network models for translation will not be required.

UCT KNOWLEDGE COOP

The UCT Knowledge Co-op builds on this tradition of social responsiveness – and aims to make it easier for community partners to access UCT’s skills, resources and professional expertise. It works by matching community groups with academic partners in a collaboration that meets the needs for research or practical support identified by the community group. The partners jointly reformulate the questions into manageable student projects. In the case of research projects, these are mostly taken up by students as projects that are conducted under the supervision of a senior academic for their dissertations. Projects address a wide range of need: from ICT support to early childhood development, prisoner rehabilitation, and sustainable community gardens.

For a list of current projects, please see visit the website:

<http://www.knowledgeteco-op.uct.ac.za/>



If you find a project that interests you, please treat it as a self-selected topic. You will need to find a supervisor willing to supervise your chosen project, and should send the MIT Coordinator both the proposed topic and your list of three potential supervisors that you would like to approach.

WALLACE CHIGONA (INFORMATION SYSTEMS)

Development of a platform to support a Community of Purpose for Cybersecurity in a School

There is an increase in the use of ICTs amongst various stakeholders of a school community to support teaching and learning as well as for social and personal use. This increase in the use of ICTs has resulted in an increase in risks associated with cybercrime. Research shows that there is limited understanding on the roles of different school stakeholders when it comes to maintaining cybersecurity. We argue that organizing the stakeholders in a community of purpose which would work together to create a safe environment would contribute to creating a safe environment. There is limited understanding on how such a community would work. The aim of this study is twofold. First, to understand what the community of purpose for cybersecurity in a school would look like. Second, to design a platform which would support such a community.

Teaching and engaging marginalized communities about data governance and AI

Many governments and organizations are engaging on how to draft legislation and policies around data governance. There is a need for citizen participation in such debates. However, citizens who are marginalized may face challenges in understanding the complex concepts of data governance. The aim of the study is to develop a prototype system which would help educate marginalized communities about the complex concepts around data governance. The system should allow the users to ask questions related to the concepts. The system could use AI to generate responses.

ZAINAB RUHWANYA (INFORMATION SYSTEMS)

Cybersecurity And Privacy Research Group (CSPR)

Our research area focuses on the socio-technical aspects of Cybersecurity, information systems security and privacy. To see the current projects list, please visit our website: <https://sit.uct.ac.za/cspr/our-research>