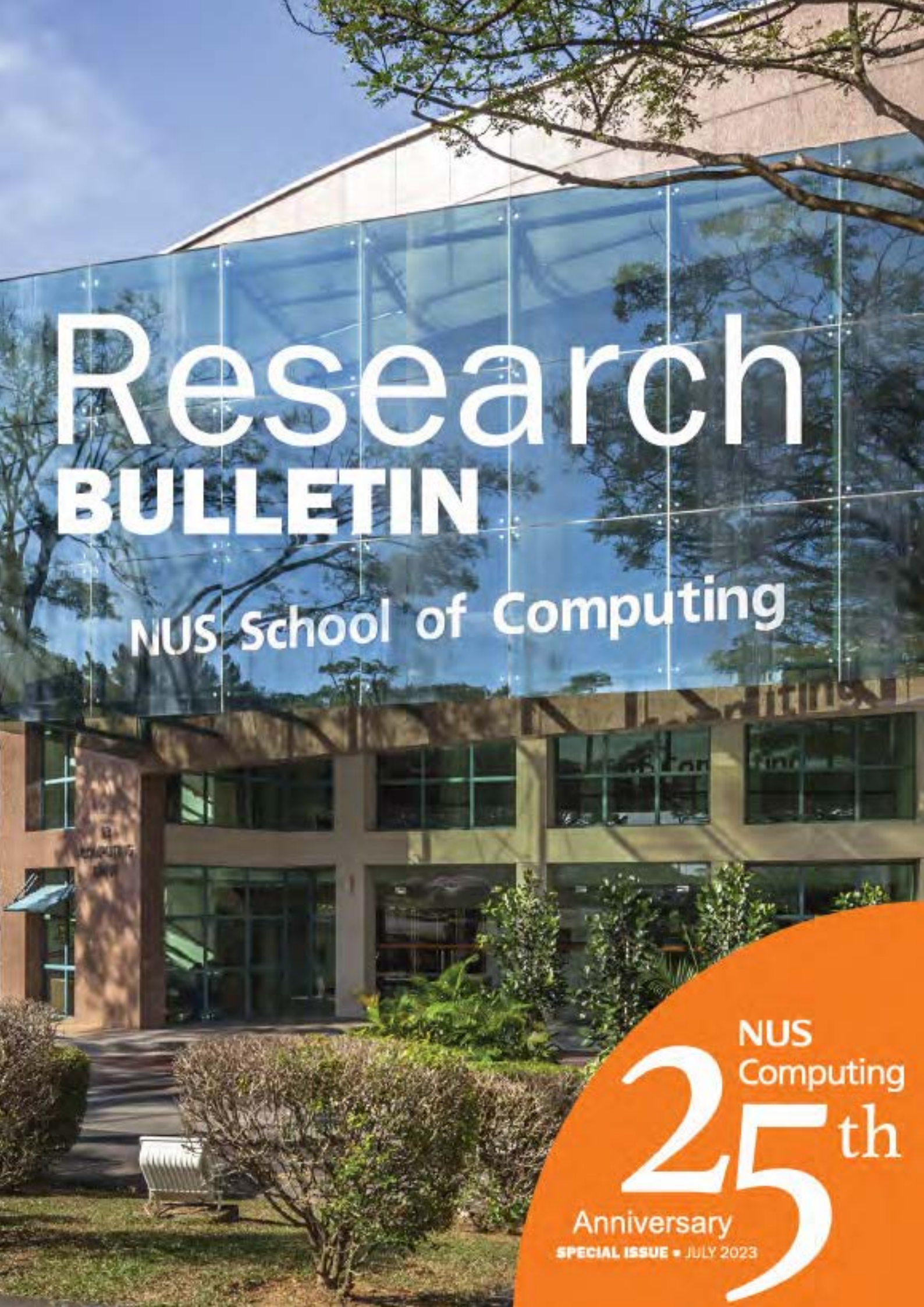


SPECIAL ISSUE • JULY 2023

Dean's Foreword

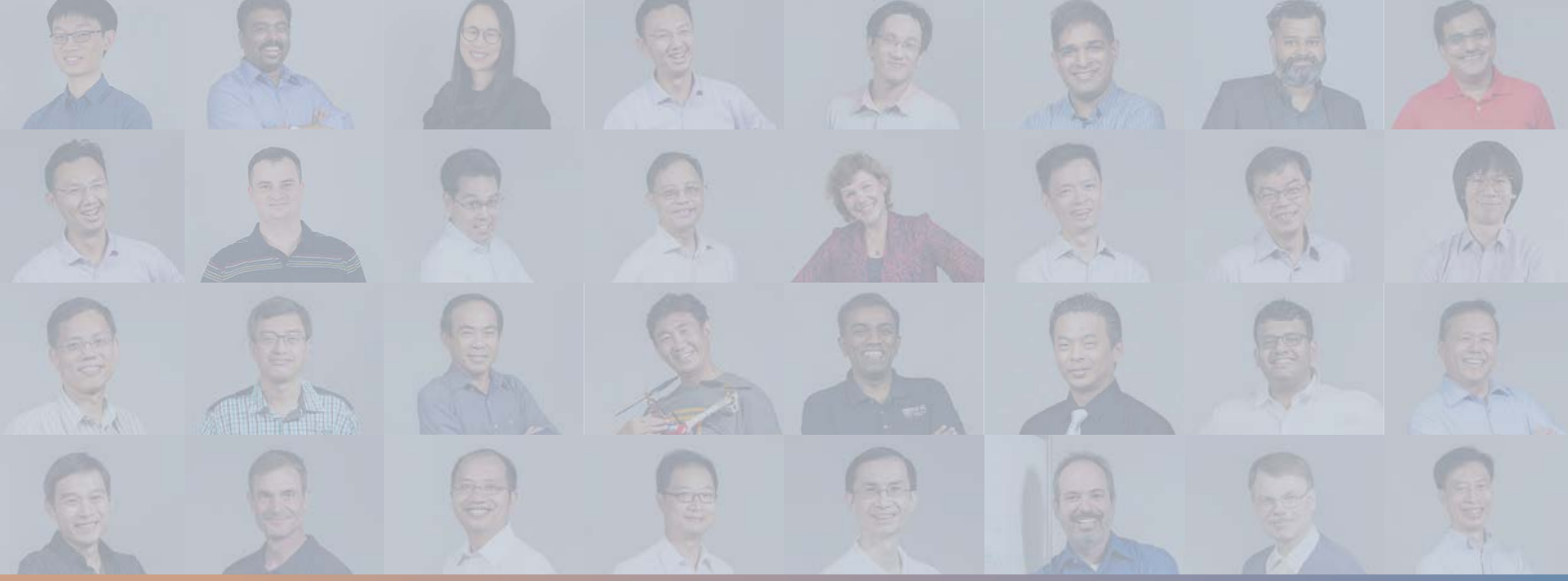
Ever since the NUS School of Computing was established 25 years ago, we have prided ourselves as a leading computing school in education and research. Our areas of research have developed from strength to strength over the years, keeping pace with fast-evolving tech trends and emerging technologies.

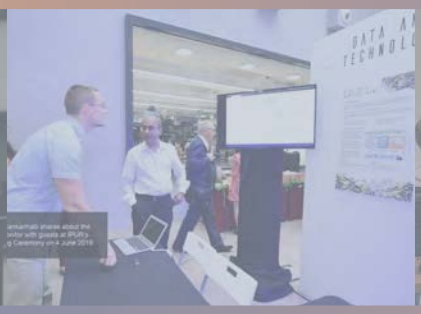
The number of research projects has registered an almost sevenfold increase since 1998, standing currently at about 350, with a research funding of over SGD230M that supports our research efforts. We have also grown our pool of research talent tremendously – from more than 40 research staff* and 200 MSc/PhD students* to 221 and 1,370 respectively today.

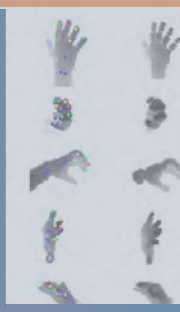
I am proud to note that the calibre of our researchers continues to make their mark globally as thought leaders. Our researchers have achieved local, regional and international recognition through fellowship awards, leadership positions and research publications. In 2022, we won many key research awards at international conferences, including 32 best paper awards. You can read more about the research articles in this bulletin.

Even as we head towards uncharted territory in tech, we remain committed to research excellence across different dimensions while innovating to develop solutions to benefit society. We hope to further deepen our research collaborations with industry and the government to fulfil this purpose, and look forward to your continual support for the next 25 years and more. Together, we can shape a better future for all.

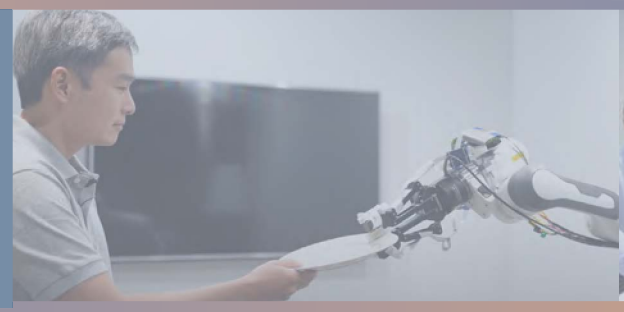
Professor Tan Kian Lee
Dean, School of Computing
Tan Sri Runme Shaw Senior Professor

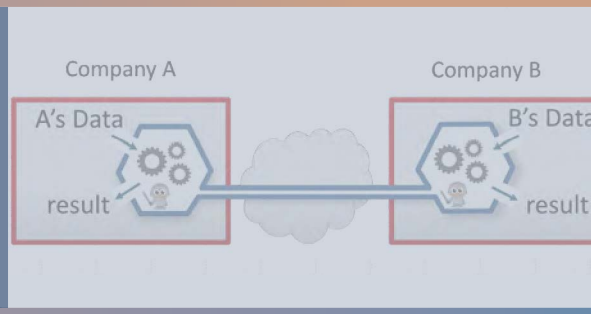


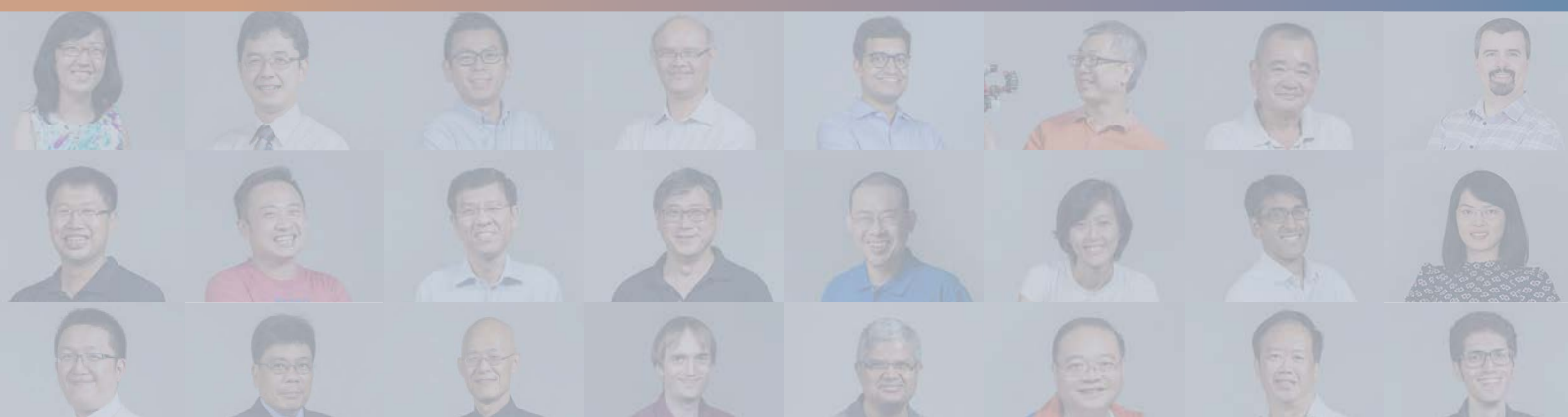





Goal Shopping List
x1 x3 x3 x1
0%?
80%?



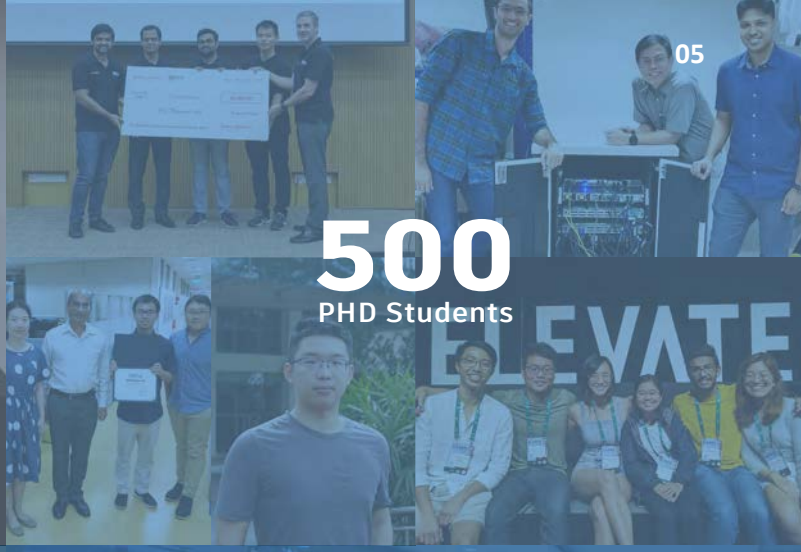

Company A
A's Data
result
Company B
B's Data
result



Research Fund

~$234M

usually spread over 3-5 years



500

PHD Students

354

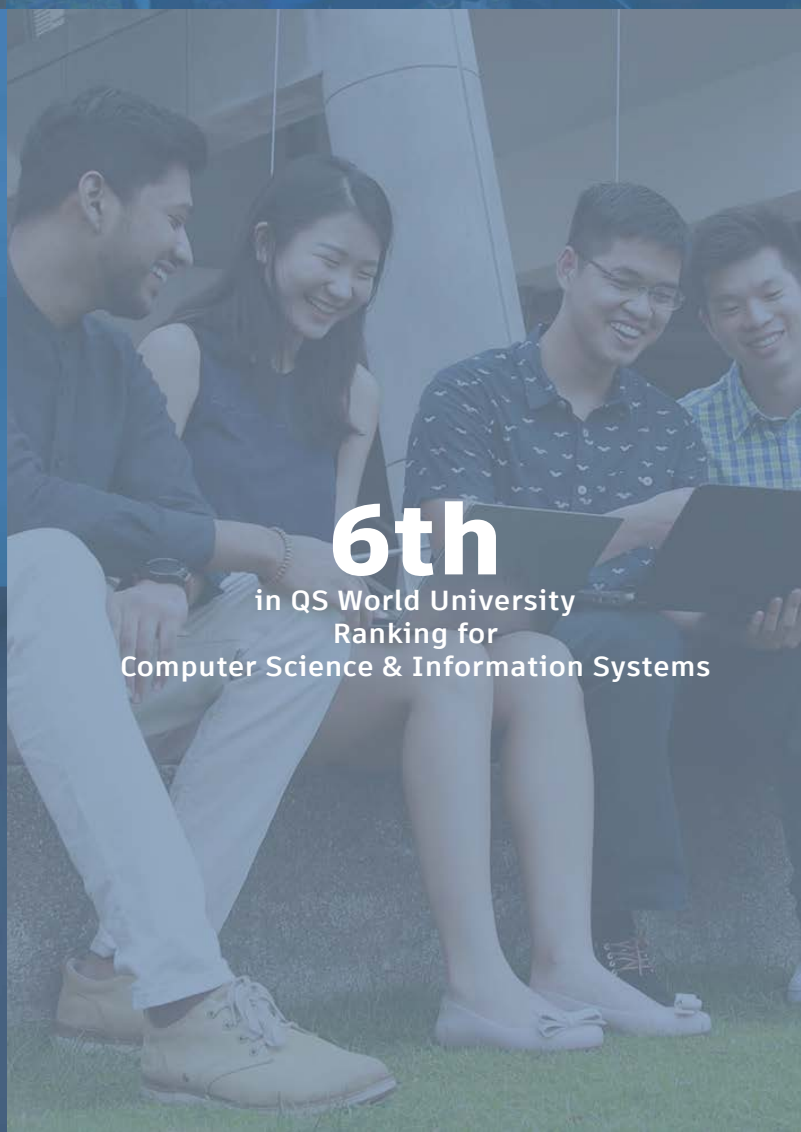
Projects

221

Research Staff

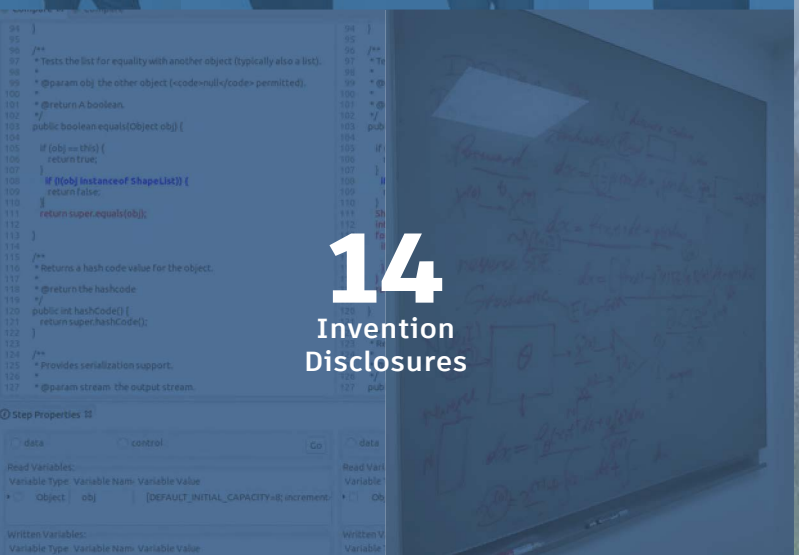
113

Principal Investigators



6th

in QS World University Ranking for Computer Science & Information Systems



14

Invention Disclosures

Department of Computer Science

The Department of Computer Science, with over 76 internationally recognised faculty members, is a leading centre for learning, teaching, and research in computer science and its applications. The department offers three undergraduate programmes in Computer Science, Computer Engineering, and Information Security, three Master's programmes in Computer Science, Artificial Intelligence, and Infocomm Security, as well a PhD programme in Computer Science.

The department's faculty members perform cutting edge research in eight research areas:

- Algorithms and Theory
- Artificial Intelligence
- Computational Biology
- Databases
- Media
- Programming Languages & Software Engineering
- Security
- Systems & Networking

Research Areas

Algorithms and Theory

Algorithms form the backbone of computer science. They determine what computers can do and how much resources are needed for a computational task. The Algorithms and Theory group aims to understand the theoretical foundations of computer science, and to use these insights to produce novel practical algorithms. This includes a wide variety of topics in theoretical computer science and its related areas. Examples include distributed algorithms, networking, cryptography, game theory, and security.

Artificial Intelligence

The term Artificial Intelligence (AI) was first coined by John McCarthy in 1955 to describe the science and engineering of making intelligent machines. Today, AI is the key technology in many novel applications, from face recognition to language translation. This research group aims to advance the development of AI by bringing together researchers across the areas of computer vision, machine learning, planning and decision making, and robotics.

Computational Biology

Present-day biomedical researchers are confronted with vast amounts of data. This has driven the development of sophisticated computational and mathematical approaches to understand how biological systems work. The computational biology research group aims to enhance the flow of information and data to enable scientists to computationally model and analyse biological systems in novel ways. Some projects undertaken by this research project include: compressed index, gene regulation, protein complex prediction, and more.

Databases

The availability of data has increased tremendously in recent years and with this comes the challenge of organising and storing data in the most efficient and effective way. This research group studies the techniques for modeling, storing, indexing, and processing different types of databases. Projects in this area include data mining, the study of data management systems, developing deep learning platforms for large datasets.

Media

From computer graphics, human computer interaction, to sound and music, the media research group seeks to understand and develop new algorithms, systems, and methods for systems to interpret and manage multiple types of data and media sources. Projects from this group include: mobile app recommendation, location analytics, social media data mining.

Programming Languages & Software Engineering

The use of software can be found in almost all aspects of life today. As the world becomes more dependent on software, ensuring that these complex systems are designed to be robust against failure becomes even more important. As software grows to become more complex, researchers and companies face the challenge of keeping their systems reliable, secure, and productive. This research group tackles these problems, focusing on improving the development, verification, and optimisation of computer programs and systems. This includes areas in program analysis and verification, paradigms in programming, parallel and distributed programming, and software validation.

Security

One of the main concerns of the world today is the issue of securing our digital lives. Without security, communications can be intercepted, modified, or spoofed and data can be commercially exploited or used for privacy invasions. The security research group studies and develops novel methods of securing computer systems against today's threats. Research areas from this group include: cryptography, network security, systems security, and more.

Systems & Networking

The Internet has become an indispensable part of modern life. Whether it is to read the latest news or communicate with other people, billions of devices are interconnected via the Internet. The Systems & Networking research group studies these large scale interconnections and how computing devices can work together to perform computations and disseminate information. Research in this area includes: cloud computing, distributed systems, mobile computing, software defined networks, and more.

Department of Information Systems & Analytics

The Department of Information Systems and Analytics has a long track record in grooming prominent leaders for the digital economy and IT workforce. It offers undergraduate programmes in Information Systems and Business Analytics, and a PhD programme in Information Systems with specialization in three areas - Economics of IS, Behavioral Science and Design Science. Graduates from this department have moved on to pursue both professional and research-based careers in a broad range of industries and job functions over the years. The department faculty members conduct research in six areas:

- Digital Transformation, Platforms & Innovation
- Healthcare Informatics
- Financial Technology (FinTech)
- Data Science & Business Analytics
- Computational Social Science
- Intelligent Systems

These six areas are the critical points of intersection among technology, policy, and people, which define the near and future digital economy.

Digital Transformation, Platforms & Innovation

Information technology (IT) is moving from a supporting role to becoming a key component of organisations' value propositions. The shift towards "digitalisation" is accelerated by the emergence of digital platforms. Additionally, innovation of services and products is enabled by platforms for crowdsourcing, crowdfunding, open data, and open innovation. This research group develops and tests models and frameworks that improve understanding and provide key insights to businesses on digital transformation efforts, and their use of platforms for innovation.

Data Science & Business Analytics

Data analytics is moving beyond simple observational insights, to predict trends and outcomes for better decision making. This research group seeks to provide tangible and novel insights from big data e.g., in retail, and social networks, as well as address fundamental big data challenges e.g., entity resolution, missing value imputation, high-dimensional clustering, transfer learning models, and causality. This informs organisations and policy makers on actionable strategies for their analytics efforts.

Healthcare Informatics

Confronting the challenges of an aging population, increased complexity in chronic and acute diseases, and exponential healthcare costs that cause divides among people, governments and healthcare service providers have looked into smart use of IT to provide safer and efficient patient-care services. For instance, integrated and personalised healthcare is possible through electronic records and telehealth applications. Also, new tools and analytic capabilities are being developed for clinical decision making, operations, and population health management. This group researches the design, management, and evaluation of healthcare IT, from which practice insights are also generated.

Computational Social Science

With more data available on every aspect of our daily lives, we are able to measure human behaviour with precision, creating an unprecedented opportunity to address longstanding questions in the social sciences. This research group uses large-scale demographic, behavioural and network data (from platform archives and field experiments) to investigate human activity and relationships, in order to answer such questions for academia and industry.

Financial Technology

Digitisation is transforming financial services (e.g., peer-to-peer lending, third-party payment) for individuals and businesses, including start-ups, SMEs, social enterprises among others. New technologies such as blockchain and cryptocurrency are providing manifold opportunities. Yet, organisations often lack understanding of how to leverage them and effectively assess the outcomes. This research group works with industry partners and government organisations to address the challenges due to such technological disruption of the financial industry.

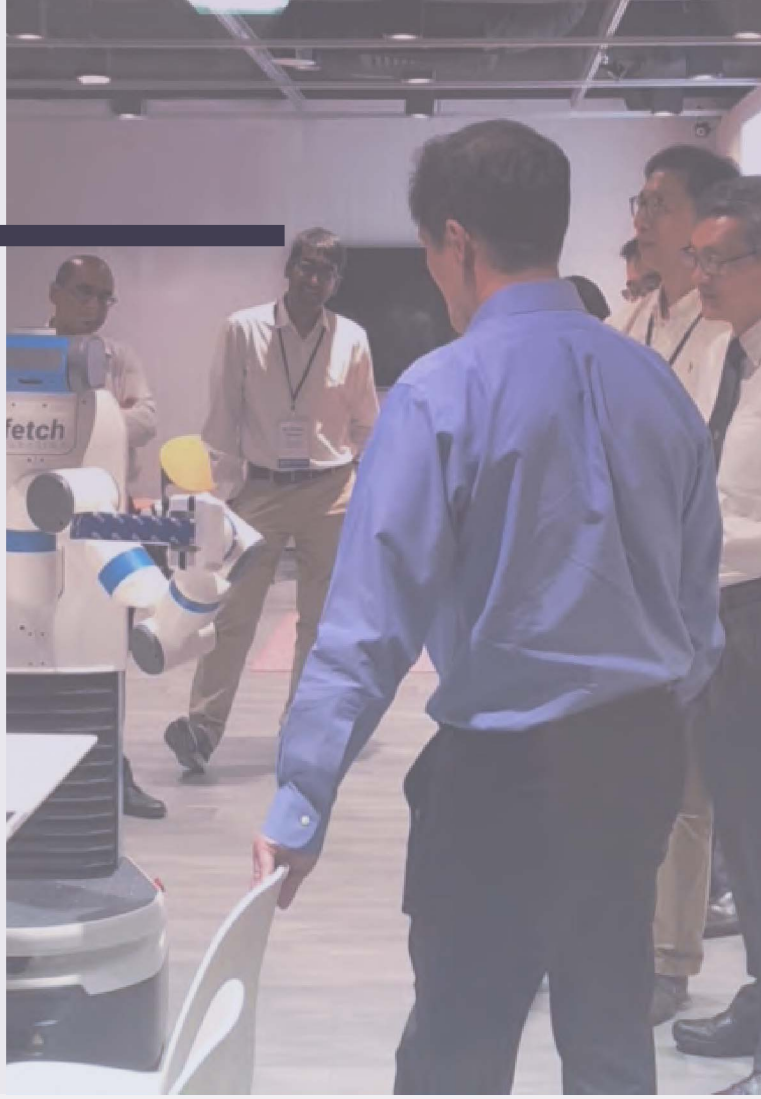
Intelligent Systems

Work automation and augmentation with artificial intelligence (AI) and intelligent systems are transforming not just organisations and industries, but the entire labour markets. At the same time, humans are learning how to interact and collaborate with smart systems and robots. This research group studies the design, use, regulation, and impacts of AI and intelligent systems on individuals, businesses, the economy and society, including its unanticipated consequences.

Research Areas

Highlights

NUS Computing is strongly committed to research excellence in all its dimensions: Searching for fundamental results and insights, developing novel computational solutions to a wide range of applications, building large-scale experimental systems and improving the well-being of society. In this research bulletin, we feature several recent efforts in Programming Languages & Software Engineering, Algorithms & Theory, Media, Systems & Networking, Data Science & Business Analytics, FinTech, and Digital Transformation, Platforms & Innovation. For research efforts in other areas, please visit: *comp.nus.edu.sg/news*



Spotting Concurrency Bugs in Software with Sampling

Presidential Young Professor Umang Mathur

In 1983, Atomic Energy of Canada Limited launched the Therac-25, a highly anticipated radiation therapy machine. However, it became infamous for causing severe accidents. Patients experienced radiation overdoses, resulting in injuries and deaths. These machines were decommissioned in February 1987, with investigations tracing their malfunction to two main causes: inadequate safety checks and bugs in the software. The Therac-25 tragedy illustrates how software errors can be devastating, says Umang Mathur, a Presidential Young Professor at NUS Computing. "In the past, in addition to loss of lives, software bugs are also a leading cause of issues such as security vulnerabilities, data corruption, crashes in software, poor performance, blackouts, and so on." Today, nearly four decades on, the problem has only intensified. "The cost of software bugs far exceeds the GDP of many nations. The real issue is that software applications are becoming increasingly complex, and ensuring software correctness is becoming more and more challenging," says Mathur, who specializes in detecting such software bugs.

Out of order

The bug in question is what's known as a *data race*. These bugs occur in software applications that run several components concurrently. "Most computing devices, including our mobile phones, run on multicore processors, meaning they comprise of small computers that run different parts of a software application in parallel," explains Mathur. "These different parts, often termed *threads or processes*, run at the same time, frequently interacting with each other to achieve a shared high-level task." Programmers often carefully choreograph the interaction between different components in concurrent software so that these components together achieve the larger task at hand. This, however, is an error-prone task and bugs such as *data races* often creep into software, often leading to failures. Over the years, computer scientists have developed ways to assist programmers to automatically detect concurrency bugs like data races by "observing the execution of the software and making an inference about whether the software exhibits a bug or not," says Mathur.

There's a high price to pay for this observation, however. "It slows down the performance of your software a lot," he says. "As a result, if I have to use these tools to detect data races, I have to think a lot to determine if it's really worthwhile to trade the performance of my software for the level of assurance given by these tools."

That is partly why most large software firms today only test for concurrency bugs (such as data races) in-house, during the initial phases of software development, before the deployment phase. But the real world can vary vastly from the controlled in-house environment. "In-house testing can often miss bugs that get triggered only when the software is actually running under heavier and more realistic workloads," says Mathur.

Sampling offers a Solution

The computer scientist began tackling the issue of data races at the end of 2021, around the time he joined NUS, after working as a researcher at Meta (then Facebook Inc.) and the University of Illinois Urbana-Champaign (UIUC), where he obtained his PhD. Last year, he and his colleagues in Denmark and the U.S. announced they had found a new way to track causality in concurrent systems — and thus detect data race bugs — in a manner that is much more efficient than existing techniques. They did this by employing a novel data structure called *'tree clocks'* to implement timestamping, an operation fundamental to many distributed and concurrent applications.

This time, Mathur has designed another method to lower the overhead of race detection. His new data race detection technique, called *Race Property Tester* or RPT for short, uses the concept of sampling to reduce the analysis cost involved in detecting data races in concurrent software. It works by examining a small proportion of the events generated during the execution of a software application, and accurately determines if the underlying application contains data races.

"The insight underpinning RPT is to treat the data race detection problem as if it were a big data problem. That is, when observing a stream of events generated during the execution of a concurrent software, instead of studying the entire stream, you only look at parts of it, which would naturally reduce your total analysis time," explains Mathur. "The best part is: the number of events that RPT needs to sample does not grow even when the execution's size keeps on growing!"

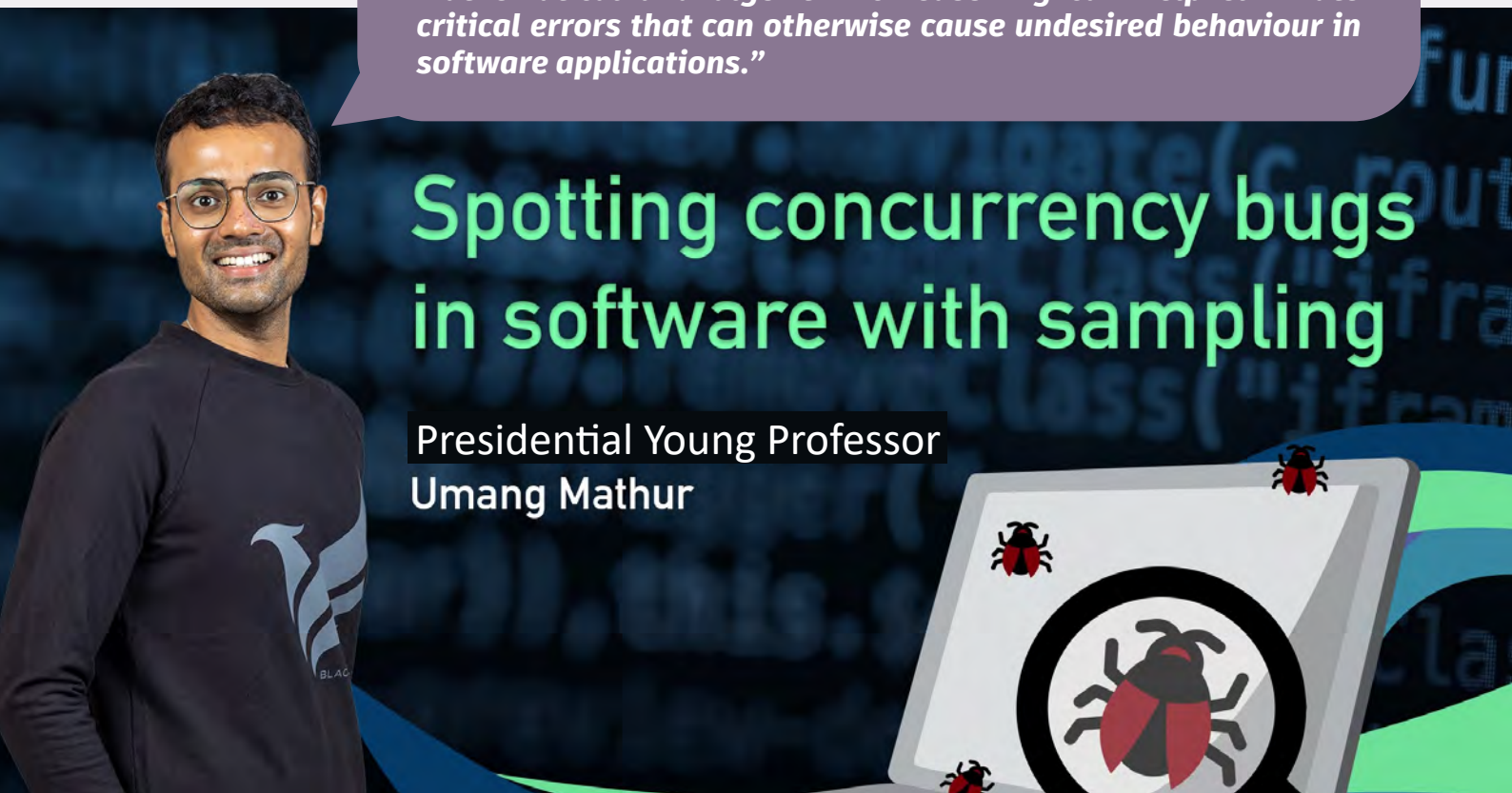
To assess RPT's performance, Mathur and his UIUC collaborators compared it against more than 140 benchmark software applications, thoroughly evaluating its effectiveness by studying factors such as run time, likelihood with which it discovers a race, whether a minimum number of bugs need to be present for the likelihood of detecting bugs to be high, and so on. "Our large-scale evaluation assured us that whatever we're doing makes sense and is indeed useful for practitioners" says Mathur. The empirical evaluation of RPT, he says, was in line with the theoretical assessment. "We found that even when software applications generate more than a billion events, RPT can detect data races with very high probability, by sampling only a very small number of events." When compared with two state-of-the-art data race detection techniques, FastTrack and Pacer, RPT demonstrated the fastest run time. Additionally, it never reported any false positives.

The researchers presented their findings at the Symposium on Principles of Programming Languages (POPL), a premier computer science conference, held in Boston, Massachusetts this January. Their work was recognized with the Distinguished Paper Award — a recognition bestowed to the top 10% of conference papers.

Mathur and his team are now working to see if they can integrate RPT into fuzz testing (an emerging first-line approach used by software firms to detect bugs) and whether they can use a similar sampling-style approach for detecting concurrency bugs other than data races. "In general, a key theme in my research is to make software more reliable and robust," says Mathur. "I think about how mathematical and algorithmic reasoning can help eliminate critical errors that can otherwise cause undesired behaviors in software applications."

https://www.comp.nus.edu.sg/news/features/2023-concurrency-bugs-umathur/

"In general, a key theme in my research is to make software more reliable and robust," says Mathur. "I think about how mathematical and algorithmic reasoning can help eliminate critical errors that can otherwise cause undesired behaviour in software applications."



Mining the Marvellous Richness of the Human Singing Voice

Associate Professor Wang Ye

Sound and music have always been a big part of Wang Ye's life, guiding him through a career that has spanned being a research engineer at Nokia in Finland to an associate professor at NUS's School of Computing. "Everybody, including myself, likes music," says Wang, who leads the Sound and Music Computing Lab.

But more specifically, Wang loves to sing. "That's probably why singing voice processing has become a key research area in my group," he says with a laugh. The singing voice is fascinating because it is unique among all other instruments, explains Wang. "It is so rich — there is nothing else that contains both lyrics and musical notation."

Parsing out these bits of information is incredibly useful. For instance, lyric retrieval — the conversion of lyrics into text, otherwise called automatic lyric transcription (ALT) — comes in handy for adding subtitles to music, indexing audio files, and aligning lyrics.

But retrieving information from a singing voice isn't easy. For a start, unlike the piano and other instruments, the pitch of a human voice is often unstable. For another, the words sung may not always be clear because singers sometimes sacrifice word stress and articulation in order to compensate for melody, tempo, and other musical aspects. Moreover, singing is often accompanied by instruments, which acts as a sort of "contamination" when you're trying to analyse the lyrics, says Wang.

The last point is especially crucial because existing lyric transcription methods tend to rely purely on audio signals as their input. "But it's very difficult to get rid of the accompanying music in the audio format," he says. "Think of rice and sand — it's very easy to mix them together, but it's very hard to separate them. It's the same with music and lyrics."

Three prongs

The solution to better lyric transcription, Wang realised, was to use other inputs — such as information captured from videos and sensors — to supplement audio signals. "***Additional modalities are very helpful," he explains. "For instance, deaf people can read lips without hearing the music. That's the trick we're borrowing here."***

The approach was unorthodox, never before tested in the world of retrieving music information. Wang, however, has always been a trailblazer of sorts: he pioneered Singapore's first and only undergraduate Sound and Music Computing course over a decade ago (it is still running today, with roughly 40 students electing to read it every semester); and in 2022, he became the first SoC professor to secure funding from the Ministry of Education's newly launched Science of Learning grant (for a project called *Singing and Listening to Improve our Natural Speaking)*.

In 2022, Wang achieved another first: when he and his students announced they had created the novel Multi-Modal Automatic Lyric Transcription (MM-ALT) system, which utilises three modalities of input — audio, visual, and signals from wearable sensors — to transcribe lyrics from singing voice.

To develop the system, the team first created a dataset they could work from. The effort was led by Wang's PhD student Ou Longshen, an award-winning violinist. Together, the researchers recruited 30 volunteers and invited them to a soundproof recording studio. There, each participant chose a song from a preset list and was filmed singing it into a microphone, which allowed the team to capture both audio and visual information. Additionally, each volunteer was fitted with an earbud that could sense how their heads, jaws, and lips moved during the recording process.

With this dataset — the first of its type to be collected — Danielle Ong, a trained linguist and part-time research assistant in Wang's lab, then worked to annotate the lyrics. Once this was completed, PhD student Gu Xiangming used his expertise in computer vision to process the videos. These three components were then integrated and used to develop the new Multi-Modal Automatic Lyric Transcription system.

Looking back, Wang says teamwork was key to their success. ***"It was a beautiful combination of different students and their respective, yet complementary, experience and talent. We put all this expertise into one pot and cooked a nice meal."***

Top marks

The results they obtained speak for themselves. When compared with audio-only lyric transcription methods, the new MM-ALT system performed demonstratively better: in scenarios where interference from musical instruments was high, the word-error rate (WER) of transcription was approximately 91% for audio-only methods. By comparison, MM-ALT's technique of using three different modalities yielded a significantly lower WER of roughly 64%.

And overall, the videos proved much more useful compared with the sensor data for boosting transcription accuracy.

With such promising results, Wang encouraged his students to write up a paper to submit to the prestigious ACM International Conference on Multimedia. To his pleasant surprise, the paper was accepted.

"It's very competitive — the acceptance rate is less than 20% and getting a paper into a top conference typically takes a lot more than two plus months my students had," he explains. "We started everything from scratch at the beginning of the semester and I didn't have much hope that they could finish the project with such stellar results."

Even more impressive, the team was bestowed the Top Paper Award at the conference last October in Lisbon, Portugal. "I'm very proud of the work my students did," says Wang.

Since then, the team has continued to work on the problem of retrieving information from singing voices — this time focusing on the other type of information it contains: musical notation, rather than lyrics. They recently submitted a paper to *IEEE Transactions on Multimedia*, which discusses how a new model they developed can extract music notes in the MIDI format based on audio-visual data analysis.

"We want to have a very natural kind of evaluation of the singing voice," Wang explains. "A singing voice has two parts: lyrics and the notes. Ideally we want to have both outputs simultaneously."

At the end of the day, he says: "I hope my research will be relevant to society and make a difference, while resonating with my students' passions."

https://www.comp.nus.edu.sg/news/features/2023-marvellous-richness-wye/

Building a Better Detector to Guard Computers Against Malicious Hardware Attacks

Assistant Professor Trevor E. Carlson

The past few years have been a mixed bag for facial recognition. In 2017, the technology stepped into the global spotlight as Apple launched the iPhone X — its first smartphone to rely on face, rather than fingerprint, scanning for authentication.

But facial recognition has also courted controversy with well-publicised studies revealing its limitations: that the software can be fooled into wrongly identifying a person using face coverings such as masks, eyeglasses, "Dazzle" makeup, and customised QR-code-like stickers. In one alarming instance, researchers took images of turtles and tweaked a couple of pixels, successfully tricking the AI system into believing the reptiles were guns.

All around the world, AI researchers followed these developments in earnest. Trevor E. Carlson, an assistant professor at NUS Computing, was one of them. He recalls: *"I started thinking to myself: can we trick a popular AI algorithm into doing something that it didn't expect to do?"*

Carlson and his team — comprising PhD students Arash Pashrashid and Ali Hajiabadi — carefully considered a handful of AI systems to study. They eventually settled on PerSpectron, a state-of-the-art tool that employs machine learning to detect malicious attacks on computer processors.

"A lot of people assume that an AI can do a good job," explains Carlson. "But there are applications that can trick the AI into thinking everything's fine, and then all of a sudden, it gets really confused and the detector can no longer uncover problematic areas with high accuracy, and the system starts leaking data." The breached data signals a loss of privacy and, if fallen into the wrong hands, can prove extremely dangerous.

Sneaking in through the side

Such attacks are known as 'side-channel attacks,' so-called because hackers use channels that unintentionally leak information to steal data from hardware systems. This indirect means to an end is one frequently adopted in science — you can't see wind, for instance, but you know it's there from the rustling of leaves in the trees or the feel of it upon your face.

Most high-performance CPUs, or central processing units, are vulnerable to side-channel attacks, says Carlson. That's because they rely on a technique called 'speculative execution' to function quickly and efficiently, by predicting the commands they're likely to receive and executing them ahead of time. "If they didn't do this, our mobile phones and laptops would be significantly slower than they are today," he says.

However, the drawback of such speculation is that it leaves behind a trail of data breadcrumbs. "The vast majority of things get cleaned up properly, but some do not," explains Carlson. "And those traces that get left behind can be detected by attackers."

It's a catch-22: CPUs can't do without speculative execution, but it can make them susceptible to side-channel attacks. Indeed, the number of such attacks — albeit theoretical for now (researchers test them on systems in a controlled manner) — has grown in recent years, with IBM, Intel, and other computer chips proving vulnerable to ominously-named attacks such as Spectre, Meltdown, and Foreshadow.

A better AI detector

Although PerSpectron is a leading AI-based, side-channel detector, Carlson's team suspected it wasn't infallible. To prove their hunch, they spent the past year developing two types of speculative side-channel attacks — which they call Expanded-Spectre and Benign-Program-Spectre — to evade the system.

When used to trick the PerSpectron detector, the researchers discovered something astonishing: the algorithm's accuracy of detecting an attack fell from 99% to 14% and 12% with Expanded-Spectre and the Benign-Program-Spectre attacks, respectively. In other words, only roughly one in every 10 attacks were being successfully identified.

"There are significant limitations with this — these AI algorithms appear to be quite fragile," says Carlson.

And so he and his team set about building a detector that could better guard against such side-channel attacks, without the need to use opaque AI algorithms that could be easily tricked. What they came up with is a system called Spectify, which they described in a new paper published in November.

Spectify works by monitoring changes to a CPU's microarchitecture — the way in which a computer is organised to carry out instructions that determines the security and performance of a system.

"*The whole idea is that we tried to break things down to the fundamental pieces," says Carlson. "All this requires very low-level knowledge of how the system functions, instead of relying on AI."*

Monitoring the microarchitecture allows one to detect the precise location of a data leak, he explains. That's because tracking changes in a CPU's microarchitecture means being "extremely detailed." Carlson likens this to keeping track of the food in your fridge in a meticulous manner: for instance, knowing that you have 15 eggs — each with its own distinctive marks or appearance — on the third rack of shelf, sitting in a transparent 5x5 plastic container that is 23 centimetres from the left wall, 32.5 centimetres from the right, and 3 centimetres from the edge.

"You need to know exactly where your eggs are before you can detect if someone has moved them," he explains. "Which egg is taken matters too."

His team designed Spectify in this manner because "it allows the system to determine the details about the data leaks before a potential attacker gets a chance to reconstruct the leaked data," thus allowing for appropriate defense measures to be deployed. "By doing so, we can restore the system to high accuracy and effectively catch all of the problems that are happening," he says. When tested against the two attacks the team created, Spectify successfully detected all attacks, with negligible false positives and zero false negatives.

Aside from this better performance, the new technique also offers up the benefits of early detection with a low performance overhead, meaning that the computer can keep running without significantly slowing down.

Carlson and his team are now exploring whether Spectify can be used to detect other types of side-channel attacks. He says: *"At the end of the day, I hope our work can lead to a new class of energy-efficient systems that can continuously and efficiently detect potential side-channel issues early."*

https://www.comp.nus.edu.sg/news/features/2023-better-detector-trevorecarlson/

So You Have a Dataset? Think About the Values It's Missing

Professor Hahn Jungpil and Associate Professor Huang Ke-Wei

Imagine that you're a book publisher gathering feedback for a new novel that your firm has recently released. Sales figures are useful, but you're keen to find out more about what people actually think of the book. So you gather Amazon-style reviews, asking respondents to rate it on a scale of one to five.

The goal was to hit 1,000 reviews but you only have 960. Do the forty missing reviews matter? Probably not, you think. 960 is still a pretty big figure. So you happily crunch the numbers and post the new book's average rating on the company's website.

But hang on, not so fast, says Jungpil Hahn, a professor at NUS Computing's Department of Information Systems and Analytics. "We tend to think that dropping a few observations here and there isn't going to matter, given that there's enough sample," he says. "But it can matter."

For instance, what if one of the reviews you omitted was from an eminent book reviewer of a leading newspaper? Or a thought leader such as Barack Obama? Or a celebrity like Oprah who has a famous book club? Their say on the new novel could hold significant sway and convert someone who's sitting on the fence into a paying customer.

The bottom line, Hahn says, is that when values are missing from a dataset, things can get problematic. This is especially pertinent because we now live in an age of big data, where voluminous complex datasets are compiled and information is extracted from them in order to guide decision making. Big data's applications are widespread: companies use it to figure out what their customers want, humanitarians use it to predict and respond to natural disasters, and physicians use it to diagnose diseases, among other applications.

"Having high-quality data is something we would all like," says Hahn, who research partly focuses on data science and business analytics. "But in the real world, this is almost never the case. You're always dealing with data imperfections."

An unseen problem

Missing values is one such imperfection. But it is a somewhat neglected issue, says Hahn, one that requires a massive rethink on the part of data scientists. He has spent the past few years trying to raise awareness of the problem, and recently published a paper on the topic, together with NUS Computing colleague Huang Ke-Wei and former PhD student Peng Jiaxu, now an assistant professor at Beijing's Central University of Finance and Economics.

The paper — which provides a comprehensive look at the problem of missing values and how to handle them — has been published in the *Information Systems Research* journal this year. "With big data, because we have so much of it, we are disillusioned to think that we can simply ignore the missingness without really thinking deeply through it," explains Hahn. "But that's very dangerous because you can get incorrect inferences if you approach your analytics in such a way."

The result? Imbalanced customer reviews for one, says Huang. Or datasets that inaccurately predict a firm's financials, or surveys that poorly reflect a company's IT resource requirements (such as how many workers are needed, what the costs are, and so on) — one of the most common types of datasets you find in information systems work, he says.

However, most companies don't currently tackle the problem of missing values head-on. Few data scientists are aware of the issue, much less know how to deal with it. "There is a standard operating procedure (SOP) of conducting research in information systems," explains Huang. ***"But somehow in that SOP, the handling of missing values is overlooked."***

"They don't even report it, and we can only guess that they just delete the data row or arbitrarily put in some values using subjective assessment," he adds. "But those measures only really work under restrictive mathematical assumptions."

Even in academia, where researchers "should be very rigorous in reporting how they use their data," people don't disclose enough information about how they handle missing values," he says.

"I think the main reason is because researchers don't really know about it," says Huang. "Which is why we've tried to increase awareness around the issue."

Awareness as the first step

To that end, the team has detailed in their paper the three main categories of missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).

It's the last two that are problematic, says Hahn. "There's no way to statistically prove whether it's type one, two, or three, but you need to think about what it's most likely to be and really think about the nature of why the values are missing."

Only after this identification can you apply the appropriate remedy, he says. The 'solutions' vary and the team offers up a handful of suggestions in their paper, including techniques such as the estimation process, maximum likelihood, and multiple imputation methods. Hahn admits that these solutions can be complex and time-consuming to use. "They reduce bias but it's also very costly, so there's a tradeoff," he says. "It's not a panacea for missing values."

His team provides the theory and mechanisms for dealing with missing values, but data scientists still have to manually apply these to their own datasets. In the future, Hahn hopes to be able to collaborate with software engineers to create a "usable piece of software" that will allow practitioners to "plug and play" their solutions.

For now, he strongly advocates data scientists to be transparent and upfront about disclosing the gaps in their datasets. And even if they don't use his team's solutions, ***"just being aware of missing values and knowing the potential biases that can arise is helpful," says Hahn. "It gives you more nuanced information."***

https://www.comp.nus.edu.sg/news/features/2023-missing-values-hjungpil-hkewei/

Is the Right-to-Repair an Overrated Battle?

Assistant Professor Chen Jin

For the most part, Henrik Huseby was an average, hardworking man — a small business owner making a modest living repairing iPhones and MacBooks in Ski, a tiny city in Norway with a population of roughly 20,000.

Huseby's ordinary life, however, turned topsy turvy in 2018 when a letter from Apple appeared out of the blue, accusing him of importing counterfeit phone screens. The charge eventually made its way to the country's Supreme Court, and culminated with Huseby being ordered to pay Apple nearly 10,800 euros in fines and court fees.

Many critics cried foul, denouncing the charge as a wider bid by the tech giant to discourage independent, third-party players such as Huseby from repairing its products. Incidences like these, they argued, were exactly why the "right to repair" movement is so critical.

The freedom to choose

The premise of the movement is simple: when you purchase something — be it a car, smartphone, or refrigerator — you should have the right to repair it by yourself or take it to a technician of your choice.

Yet, these options are often unfeasible because "big tech firms usually have a monopoly on the repair market," says Chen Jin, an assistant professor at NUS Computing. Manufacturers employ all sorts of tricks to ensure customers return to them for repairs, for instance withholding spare parts, soldering devices together unnecessarily, or disabling custom software. They argue that keeping repair largely in-house helps maintain product safety while protecting intellectual property rights.

But in the past decade, the right to repair movement has gathered steam, calling on governments to pass legislation to combat this, by requiring manufacturers to share repair information, diagnostic tools, and replacement parts.

"Right to repair activists believe that manufacturers having a monopoly on the repair market is bad, and that more competition will benefit consumers," explains Jin. The environment could benefit too, with less e-waste generated as people would be more encouraged to mend broken gadgets rather than simply purchase new ones — something that steep repair costs currently deter.

"Conventional wisdom suggests that the right to repair benefits consumers and reduces environmental impact," says Jin.

But would it really be a clear-cut win-win scenario? Jin wasn't so sure. Passing such legislation would understandably hurt manufacturers' profits, but it was unclear how companies would try to mitigate these losses. "How would they adjust prices,especially that of new products? And what are the welfare and environmental implications of such changes?"

"Any assessment of right-to-repair is incomplete without incorporating this pricing perspective," says Jin. "Yet it's largely missing in today's policy debates."

To fill this knowledge gap, Jin — together with PhD student Cungen Zhu and fellow researcher Luyi Yang, an assistant

professor of operations and IT management at the University of California, Berkeley's Haas School of Business — set out to create a model to analyze the implications of right-to-repair.

Taking a second glance

The results they obtained are illuminating. *"Our research challenges conventional wisdom," says Jin. "What we're saying is that those intuitive predictions may be true, but hold on a second as it's not always the case — under some conditions, the consumer and environment can be worse off if a right-to-repair bill is passed."*

The key, he says, is to look at how much it costs to produce the good in question. If something is relatively cheap to make, for example smartphones and microwaves, the new model predicts that a right-to-repair bill will likely see manufacturers lowering the price of their new products.

"Doing so disincentivises repair and resale," says Jin. "The firm wants consumers to buy the new product instead of fixing a used product." Slashing fees would also help firms avoid old products cannibalising the sales of new ones. While this benefits buyers, the environment could lose out with more e-waste generated.

In contrast, when production costs are high, the firm's new goods inevitably come with a steep price tag. To overcome this and stimulate demand, manufacturers are likely to offer free repair services to whet customers' appetites. This is because of the 'value enhancement effect,' i.e., the repair service could make the products last longer and hence, increase the consumers' valuation of the product.

Since repair is offered free of charge, a right-to-repair legislation — along with its resulting lower repair costs — is unlikely to make much of a difference.

In instances where production costs are intermediate, Jin's model predicts that the outcome is a combination of the above effects. When right-to-repair is introduced and independent repair costs start to fall, a manufacturer is likely to lower the prices of their new products in order to entice customers away from the repair option (similar to what happens in the low production cost scenario). This benefits customers but may be detrimental to the environment.

But as independent repair costs continue to fall beyond a certain point, this price-slashing option is no longer attractive to manufacturers, he says. In response, manufacturers are likely to switch tacks and raise new product prices while simultaneously offering free repairs (i.e. the high-production cost strategy). "All told, if the use impact of a product is the main contributer to the environmental impact, then a right-to-repair bill in this scenario can create a lose-lose-lose situation that compromises manufacturer profit, reduces consumer surplus, and exacerbates the environmental impact," says Jin. "That's despite repair being made easier and more affordable."

His team's findings were published earlier this year, as a preprint in the journal Management Science. Next up, the researchers plan to modify their model to be able to analyse a non-monopolist market. *"We want to look at the effect of right-to-repair on price competition, for example between Apple and Samsung in the cell phone or electronic device market,"* explains Jin.

To him, the topic remains fascinating, not to mention timely — in 2019, Apple launched a new repair programme offering more third-party shops repair tools, parts, and information; while in 2021, the United States, European Union, and British governments introduced right-to-repair laws.

Jin concludes: *"Our results tell a cautionary tale and urge legislative authorities to factor in the inextricable link between repair and product markets when they assess the right-to-repair."*

https://www.comp.nus.edu.sg/news/features/2022-right-to-repair-chen-jin/

Research Centres

2022

SIA-NUS Digital Aviation Corporate Laboratory

In January 2022, SIA and NUS launched the SIA-NUS Digital Aviation Corporate Laboratory which aims to create and commercialise innovative technologies that could accelerate the digital transformation of Singapore's aviation sector, redefine the air travel experience and ensure safety and security in air travel. This will be achieved by leveraging NUS's world class deep tech and multi-disciplinary research expertise across artificial intelligence, machine learning, data science, operations research and analytics, optimisation, automation, sleep studies and design to develop digital technologies. These technologies will be evaluated and potentially be adopted for use by SIA.

siacorplab.nus.edu.sg
Work Package 3 – Employee Wellness
Led by Professor Teo Hock Hai

2021

NUS-NCS Joint Laboratory for Cyber Security

NUS and NCS Pte Ltd have established a joint research lab that is hosted in NUS to conduct research, develop capabilities and innovative digital solutions to protect individuals, businesses and public agencies in Singapore from a wide range of cyber threats. The joint lab is governed by a Management Committee comprising members from NUS and NCS to conduct research in three broad areas of cyber security having strategic relevance to NCS's business: AI Security; Cyber-Physical System Security; and Data Security and Privacy.

nus-ncs.nus.edu.sg
Led by Professor Chan Mun Choon

2020

Singapore Blockchain Innovation Programme

Anchored in NUS, SBIP aims to align blockchain technology research with the needs of the industry, to facilitate the development, commercialisation and adoption of wider real-world applications.

sbip.sg
Led by Professor Ooi Beng Chin

2019

NUS AI Lab

NUSAiL is hosted within the NUS Computing with members from the school as well as affiliated members from other faculties and organisations. It aims to be a centre of excellence in AI research, education, and practice. The research covers theory, machine learning, reasoning, optimisation, decision making and planning, modelling and representation as well as computer vision and natural language processing.

ai.nus.edu.sg
Led by Professor Leong Tze Yun

CRYSTAL Centre

2018

Located in NUS Computing, the CRYSTAL (Cryptocurrency Strategy, Techniques, and Algorithms) Centre is an academic research laboratory and think-tank that aims to provide scientific clarity in the blockchain and cryptocurrency space. The centre conducts research on current blockchain and cryptocurrency issues like scalable consensus protocols, verification and testing techniques, and privacy-preserving computation.

crystal.comp.nus.edu.sg
Led by Associate Professor Prateek Saxena

NUS-Tsinghua-Southampton Centre for Extreme Search

2016

NExT++ is a joint research centre established by NUS, Tsinghua University of China, and the University of Southampton, United Kingdom. Co-hosted in NUS Computing and the NUS Smart System Institute, NexT++ carries out research into big unstructured data analytics and its paradigm changes. The centre focuses on research in three main areas: multi-modal multi-source data analytics, video object relations, and recommendation systems.

nextcenter.org
Led by Professor Chua Tat Seng

Advanced Robotics Centre

2013

ARC is an interdisciplinary research centre, established by the NUS Computing and NUS Engineering, to lead and support robotics research in NUS and Singapore. The centre focuses on developing human-centred collaborative robotics, with the goal of refining the scientific foundations, technologies, and integrated platforms to enable human-robot interaction and collaboration.

arc.nus.edu.sg
Led by Professor David Hsu

National Initiatives

2021

DesCartes

The five-year DesCartes Programme, established in October 2021, which is funded by the National Research Foundation, Singapore (NRF), aims to enhance real-time decision-making in urban-critical systems, with a focus on individuals and society at large.

One of the Co-Directors of the DesCartes Programme is Professor Abhik Roychoudhury from NUS Computing, who will focus on the foundations of intelligent computing and its translation in the different case studies, for trustworthy integrated decision making in smart cities. He and his NUS team will validate techniques involving formal methods, intelligent control, smart data, and human-AI collaboration on a large number of industrial case studies including electricity network, smart building and future transportation networks. They will also work with the other partners to ensure that the foundational concepts are translated via industrial collaboration.

cnrsatcreate.cnrs.fr/descartes/

Work Package 1 —
Led by Professors Abhik Roychoudhury & Blaise Genest

Work Package 4 —
Led by Professor Christophe Jouffrais & Associate Professor Ooi Wei Tsang

2020

Centre for Trusted Internet and Community

CTIC is a university-level research centre dedicated to the interdisciplinary study of the Internet and its implications on the society of the future. It adopts a unique approach of integrating three different perspectives- technology, human and policy- to develop a set of insights, tools, policies and best practices around the use of the Internet to promote digital well-being and responsible public discourse.

ctic.nus.edu.sg

Led by Professor Lee Mong Li, Janice

2018

NUS Centre for Research in Privacy Technologies

Located in NUS Computing, N-CRiPT is funded by NRF and administered by IMDA. Working 'Towards a privacy-aware Smart Nation', the centre's goal is to develop privacy-preserving technologies to protect privacy at an individual and organizational level in a holistic manner- with focus on, but not limited to, unstructured data - along the whole data life cycle.

ncript.comp.nus.edu.sg

Led by Professor Mohan Kankanhalli

2017

Singapore Data Science Consortium

The SDSC was established by NUS, NTU, the Singapore Management University, and the Agency for Science, Technology and Research (A*STAR), to empower Singapore to harness the power of data science. The consortium helps industry partners to access the latest data science technologies, applications, and expertise from academia to create innovative solutions for real-world challenges.

sdsc.sg

Led by Professor Tan Kian Lee

AI Singapore

2017

AI Singapore is a national AI programme launched by NRF to develop national capabilities in artificial intelligence, thereby creating social and economic impact, growing local talent and building an AI ecosystem. AISG is hosted by NUS and brings together research institutions and the vibrant ecosystem of AI start-ups and companies to conduct use-inspired research, and develop the talent to power Singapore's AI efforts.

aisingapore.org

Led by Professors Ho Teck Hua, Mohan Kankanhalli & Chen Tsuhan

Institute of Data Science

2016

IDS is a university-level research institute established to develop integrated data science solutions and to nurture the next generation of data scientists for Singapore's Smart Nation initiative. Within the span of 5 years, IDS has successfully built up a core team with a portfolio of transdisciplinary and translational data science projects and strategic partners. IDS has been ahead of the curve to become a world-class institute in data science research and innovation, establishing the first AI industry research lab in NUS with Grab in 2018. The institute focuses on establishing a core research agenda for data science, thematic flagship programmes, international profile and networks; and developing a cross-cutting data science platform.

ids.nus.edu.sg

Led by Professor Wynne Hsu

National Cybersecurity Research & Development Laboratory

2015

Many communities and informal businesses in developing nations typically lack proper service from existing technology vendors and telecommunication companies. To meet this need, COSMIC aims to empower underserved communities through social media innovations to improve the way they live, work, and play. The centre is a collaborative initiative between NUS, the Nanyang Technological University (NTU), and the Indian Institute of Technology Bombay. Projects from this centre include an intelligent pest-control mobile app solution for farmers and creating a musical habilitation framework for children with post cochlear implantation.

ncl.sg

Led by Associate Professors Chang Ee-Chien & Liang Zhenkai

Smart Systems Institute

2007

Led by NUS Computing, SSI is an interdisciplinary research institute and an experimental "playground" for human-centered AI engineering. The research at SSI focuses on AI and robotics technologies for interaction design and embodiment, with the aim to enable situated AI assistance for every human to work, play, and learn.

ssi.nus.edu.sg

Led by Professor David Hsu

Image & Pervasive Access Lab

1998

IPAL is a Franco-Singaporean International Research Laboratory formed via partnerships between the French National Center for Scientific Research (CNRS), the Agency for Science, Technology, and Research (A*STAR) and the National University of Singapore (NUS). First established in 1998 as a special CNRS overseas laboratory, IPAL serves as a framework for collaborations between NUS, A*STAR, and French researchers. The IPAL partnership is joined by French institutions over the years (Joseph Fourier University (later as Grenoble Alpes University), Institut Mines-Télécom, University Pierre et Marie Curie (later as Sorbonne University), and National Polytechnic Institute of Toulouse).

ipal.cnrs.fr

Led by Associate Professors Christophe Jouffrais, Ooi Wei Tsang & Dr Lim Joo Hwee

nus.edu/computing