Report on the V-KEMS/IMI Virtual Study Group: *Mathematics for Justice*

20^{th} - 22^{nd} November 2023

Prof. Chris Budd OBE, Dept. of Mathematical Sciences, University of Bath

1. Executive summary

The Virtual Study Group (VSG) on Mathematics For Justice was conducted online on 20^{th}-22^{nd} November 2023. It was organised by a steering committee chaired by Chris Budd and comprising a mixture of academics and members of the justice system (in particular the Ministry for Justice). The meeting was facilitated by the Bath Institute for Mathematical Innovation (IMI) in collaboration with V-KEMS, and with further administrative support from the ICMS. The meeting was supported financially by ICMS (through the Maths For Humanity Scheme) and the KE-Hub.

The aim of the meeting was to provide a framework for (mathematical) collaborations between mathematicians and members of the justice system, with the idea of identifying key problem areas which would both benefit from a mathematical approach, would lead to new mathematical research and ideas, and would generate new collaborations. To do this three thematic areas were addressed in (i) flow through the justice system (ii) statistical evidence of bias in the justice system (iii) the role of AI in the justice system. Great progress was made on each during the VSG during which mathematical tools were found to be very effective in explaining a number issues and making better sense of the judicial data. Substantive reports on each thematic area are being written and will be posted on the V-KEMS website. Short summaries on the progress in each theme are given in this report. In addition a popular article based on the results obtained in the meeting will shortly appear in the online PLUS-Maths magazine.

The meeting was attended by a mixture of around 40 mathematicians, statisticians, operational research researchers, and AI experts from around the world, together with members of the Ministry of Justice, barristers, and Police Scotland.

Plans are already in place for a one day follow up meeting for the VSG. In addition the Newton Gateway has just hosted another three day VSG on: Food Security.

2. Background: Knowledge Exchange, V-KEMS and Virtual study groups

Mathematical Knowledge Exchange can best be described as the mutual exchange of mathematical knowledge between mathematicians, other disciplines, and end users, to the mutual benefit of all. New mathematical ideas can be of direct benefit to end users Working on challenging problems in (for example) industry can directly lead to new mathematical

developments and research areas. In addition great mathematics is often done outside of universities, and academic mathematicians have much to learn from end-users, and end-users from academic mathematicians.

The UK has a long tradition of such Knowledge Exchange. For the last 50 years this activity has been greatly stimulated by the operation of week long 'Study Groups' in which industrialists and mathematicians meet on equal terms for brainstorming sessions centred on applying novel mathematical ideas to challenging industrial problems. The study group concept has proved remarkably successful in establishing links between academia and industry, finding mathematical solutions to practical problems, generating new mathematical research ideas, furnishing excellent case studies for teaching, and training staff and PhD students. See http://www.cms.ac.uk/wp/wp-content/uploads/2019/11/StudyGroupReport.pdf for a data informed review of the impact of study groups. The study group concept has now spread all over the world and the UK study groups are now part of the more general programme of European Study Groups (ESGI).

Early in the COVID-19 pandemic it was realised that the Study Group approach afforded a very effective way of bringing mathematical expertise to bear on the problems posed by the emergency. Accordingly the Virtual Forum for Knowledge Exchange in the Mathematical Sciences (V-KEMS) was founded to organise and run frequent online virtual study groups (VSGs) focused on applying mathematics to COVID-19 related problems. Each VSG lasted for three days and was followed shortly afterwards by a careful report. The VSGs proved remarkably effective, and the results from the VSGs were used to help inform government decisions. V-KEMS was awarded the Praxis-Auril KE prize for this work and members received commendation from the Government's Chief Scientific Advisor and Chief Medical Officer.

The VSG approach proved very effective in bringing a wider variety of mathematicians and end users together than was possible in a 'traditional' study group, allowing a richer variety of both problems and mathematical approaches to be explored. (The mathematical techniques include modelling, data science, statistics, operational research, network theory, dynamical systems, and machine learning). Following the end of the pandemic it was felt by the organising committee that V-KEMS should continue to run VSGs in parallel to the face to face study groups. To make them distinct, and to take advantage of the flexibility and universality of the VSG concept, it was agreed that the focus should be on the application and development of mathematics to problems of a more social science nature. This resonates well with the philosophy of the 'Maths for humanity' programme of the ICMS

The initial plan agreed by V-KEMS was for three VSGs in 'Maths for Justice', 'Food security', and 'The Future of the High Street'. The first of these, organised by V-KEMS and the Bath Institute for Mathematical Innovation, and supported by ICMS and the KE-HUB is the subject of this report.

3. Overall structure and operation of the Maths for Justice VSG

3.1 The steering committee and overall planning

To organise the VSG the following organising team was put together which brought together skills in mathematical modelling, statistics and data science, operational research, and AI, together with domain experts from the justice system including the Ministry for Justice (MOJ)

- Chris Budd OBE mascjb@bath.ac.uk (Chair. IMI Bath)

- Henryk Faas henryk.faas@justice.gov.uk (MOJ)
- Jane Hutton j.l.hutton@warwick.ac.uk (Warwick)
- Lauren Hyndman lauren.hyndman@icms.org.uk (ICMS)
- Jose Pina Sanchez j.pinasanchez@leeds.ac.uk (Leeds)
- Lizzie Tiarks elizabeth.tiarks@northumbria.ac.uk (Northumbria)
- Julio Trecenti julioazt@insper.edu.br (Insper, Brazil)
- Machi Tseloni andromachi.tseloni@ntu.ac.uk (MOJ/NTU)
- Myla Watts pysmad@bath.ac.uk (Administrator, IMI Bath)

The team met regularly (online)to identify the appropriate themes, solicit the correct data, and to plan the operation of the VSG. Chris Budd also serves on the Executive Board of the KE-Hub and kept them fully updated. The overall organisation of the group, and of the VSG itself was coordinated by the IMI Administrator Myla Watts..

Much of the operation of the VSG during the three days, and the writing of the reports, was done by the 'facilitators': Tosin Babasola (Bath), Sam Kamperis (Oxford Brookes), William Lee (Huddersfield), Ann Smith (Huddersfield), Tina Zhou (Bath).

3.2 Theme organisation

The research was done in the three themes of A. Modelling throughput through the justice system, B. Statistical analysis of bias in the Justice System, C: AI for the justice system

Each theme had an expert theme lead who was a member of the steering committee and had the job of stating the problem, producing a problem description and finding suitable literature/reports. These were A: Henryk Faas, B: Jane Hutton, C: Lizzie Tiarks

There was also an overall Data Expert, Prof Machi Tseloni, who expertly coordinated the (huge and thorny) issues of making justice related data freely available to the group.

As stated above, each theme also had a ‘facilitator’ whose job was to make sure that the team working on the theme had a clear direction and produced effective work. The facilitators also took the lead in writing the reports on each theme**.**

A Google drive was created for each theme into which were placed the problem descriptions various examples background literature and the data described above. In particular, data was obtained from the Brazil justice system and (carefully created synthetic data) from the MoJ /ADR UK Data First Programme (for which we would like to record our very great thanks). Public domain data on stop and search for 2017 to 2022 and on sentencing was also used.

During the operation of the VSG the attendees were free to choose which Theme they worked on, and it was the job of the facilitator to ensure that each Theme had a smoothly operating team which was able to deliver results.

3.3 Programme

On the first day of the VSG there were short presentations to the whole group on the three themes by the theme experts followed by a general discussion with Q+A. After a brief introduction to the software tools, there was then a brief period during which the teams for each Theme were formed (by self selection). These teams went into Zoom Breakout rooms for facilitated brainstorming sessions. At the end of each day the whole group reconvened with the teams reporting on their progress, with a substantial final presentation at the end of the workshop. On the afternoon of the second day there was also a presentation by Harry Schone of Police Scotland, on the uses of mathematics and statistics in operational police work.

3.4 Software tools

It is not easy to do mathematics collaboratively online, but a series of software tools, the use of which was developed during the pandemic, makes it easier. Chief amongst these is HackMD which is a mark down version of LaTeX. This allows a rapid sharing of mathematical ideas, data and figures and also facilitates the group working together to build up a set of notes and record of the event. The MURAL platform is excellent for developing the teams and as good way of coordinating the whole event. Google Drive serves as a repository for data, presentations and useful literature. Finally Overleaf is an excellent tool for producing the reports after the meeting.

3.4 Operation of the study group

About 40 mathematicians and members of the judicial system attended the VSG over the three days. This meant that on average about ten people were working on each problem at any one time. The online nature of the VSG meant that we had attendees from all over the world in multiple time zones (and with no Visa restrictions!). A few more people turned up for the special presentation by Police Scotland. The online format also meant that the teams could contribute to work on the problems outside of the hours of the operation of the VSG, and

many did! It also meant that those with teaching or caring commitments could easily drop in and out of the team discussions as they developed. A list of the 55 people who registered for the VSG is enclosed.

3.5 Reporting

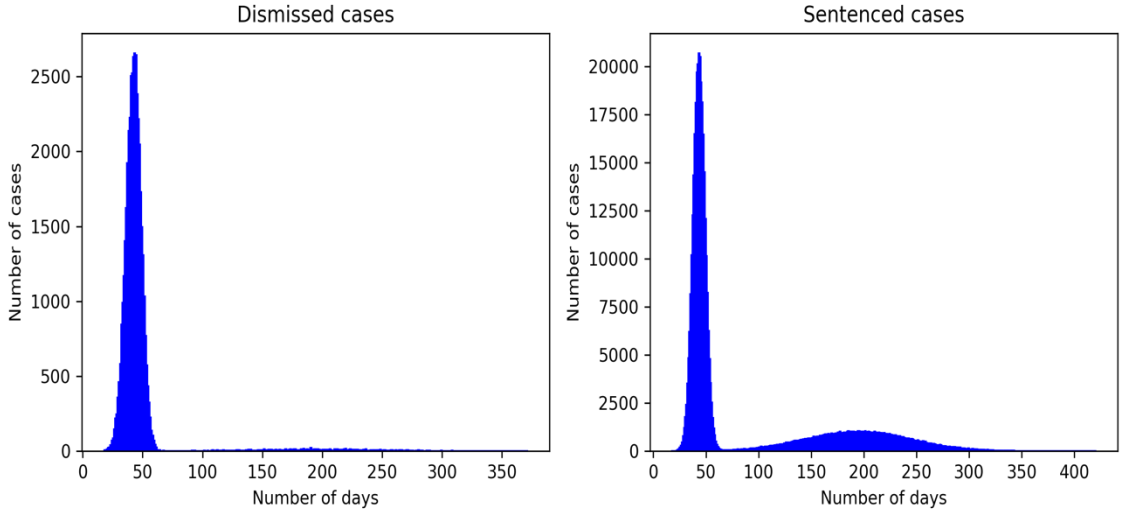
Short reports on the progress in each Theme (based on the final presentations on the third day) were provided by the Facilitators in the month after the VSG and are summarised below. The longer and more considered reports are being written (in Overleaf) and are being coordinated by Lauren Hyndman of the ICMS. They are all in an advanced state of writing and are expected to be completed in February (3 months after the meeting).

4. Summary of the research done at the VSG

The research was done in the three themes of A: Modelling throughput, B: Statistical evidence for bias, C: The role of AI in the Judicial System

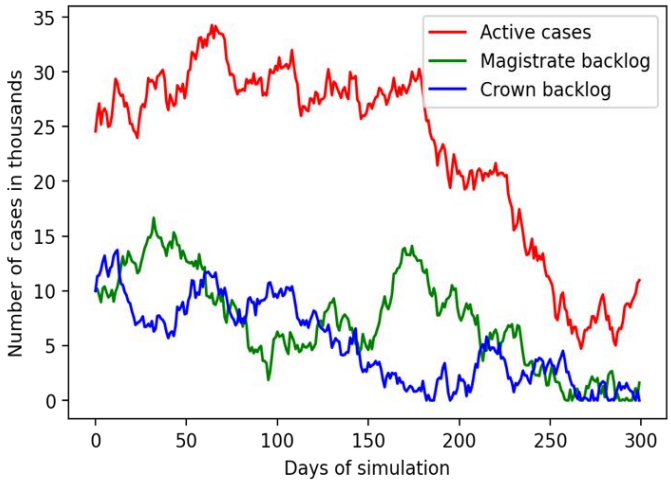
Theme A

The aim of this theme was to predict the case volume and case duration in the judicial system with a focus on the case characteristics and also the identification of the main factors that impact duration of court case. To do this the group examined the data from the national statistics for the Criminal court statistics quarterly: April to June 2023 and in particular considered the crown and magistrate court case characteristics. The group then developed both an agent based and a distribution-based model to determine the proportion of the magistrate and crown court backlog cases from the total active cases.



The agent-based model functioned by sequentially processing cases through the justice system and the group monitored the duration of each case. The model first examined the

dismissal or sentencing cases; then it proceeds to the next case. The distribution-based model was formulated to partition the system into three stacks: active, magistrate, and crown. After each time step, a random number of cases was transferred from one stack to another. There are initial periodic peaks in the number of active cases, and these peaks are sometimes accompanied by increases in magistrate and crown court backlogs.
There seemed from the model to be an inverse relationship between the crown and magistrate backlogs.



Analysing the differences and patterns in these backlogs provided useful insights into the legal processes.

For follow up the group recommended the use of additional information, such as the types of cases being filed, the reasons for case dismissals, and the average time it takes to resolve cases might be useful in identifying model parameters.

The model derived simulated a simple version of a justice system from (Summoned, Magistrates, Crown court, Sentencing). A more robust simulation could be conducted to fully implement the flow chart of the main court processes for criminal cases provided by the Ministry of Justice in the Guide to criminal court statistics report.

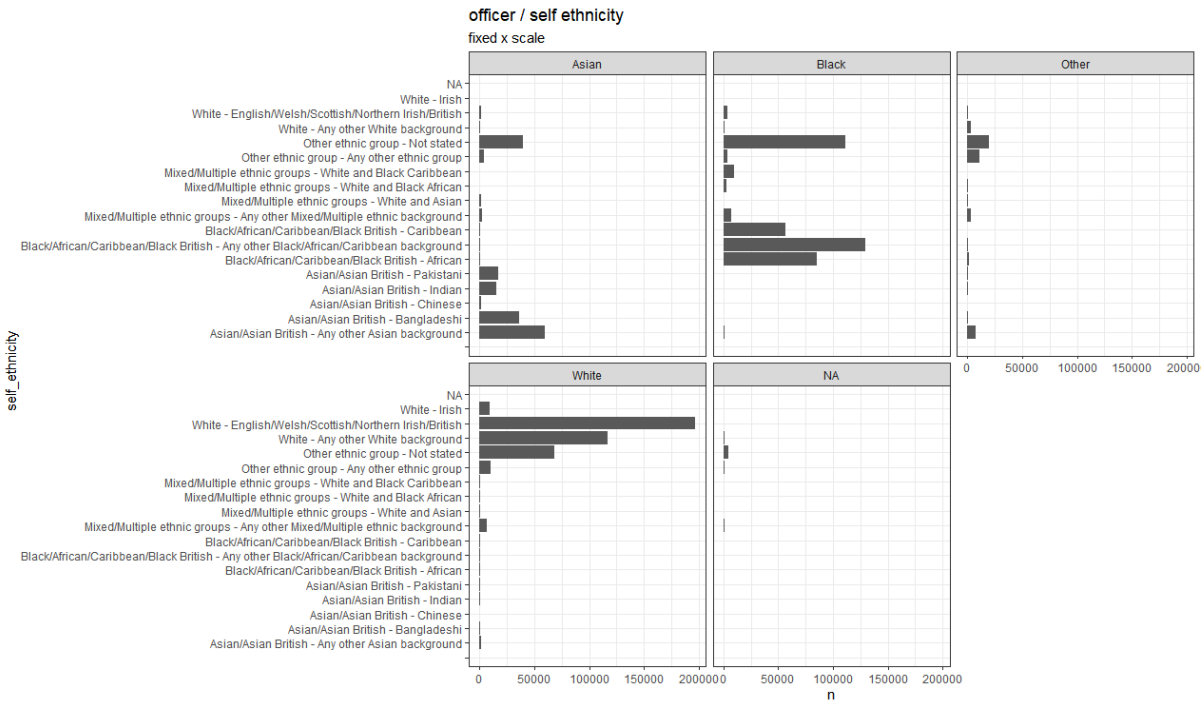
Theme B

Confidence in the justice system relies on it being both impartial and being seen to be impartial. It is therefore troubling that that there are well documented disparities between men and women and between ethnic groups at various stages of the criminal justice process. These disparities don't necessarily indicate bias or discrimination, as they might stem from correlations with legitimate factors. A perceived bias towards males in sentencing could, for instance, reflect a correlation where thefts involving force or threats (more often committed by males) result in a higher probability of custodial sentences. In this hypothetical case the true causal link here would be between use of force or threats and the sentencing outcome, rather than gender. Theme B focussed on statistical analysis of existing data to (1) better

understand the existing disparities, (2) investigate what information is available in the data that might be useful in drawing a distinction between causation (discrimination) and association, (3) consider what statistical tools could be used to quantitatively investigate these relationships.

Two datasets were investigated, looking at two different ends of the justice system process. The first dataset describes the Metropolitan Police Force's stop and search practices from an entry point into the criminal justice system. The second dataset describes sentencing, the final stage of that system, focussing on shoplifting offences dealt with by a magistrate's court.

The stop and search dataset showed a rich texture of information including temporal and geographic information as well as details of the suspected offence and the outcome of the search. Disparities in outcomes were evident in the dataset, for instance men were more likely to be arrested than women. A particularly interesting window on the thinking of the officers involved was provided by considering the difference between self-defined ethnicity and officer defined ethnicity. Plots summarising this relationship are shown below. Graphical statistics models will be important in exploring the complex relations between the purpose and outcome of searches and demographic factors.



The sentencing dataset included 36 factors describing whether a custodial sentence was applied, gender and culpability, harm and any aggravating or mitigating circumstances. An idea for investigating bias by generating synthetic data sets from which the possibility of bias had been removed allowing a bootstrap style significance test was described, which may warrant further investigation. Logistic regression was used to model the sentencing decision, with a sparse model performing as well as a full model suggesting the potential for a simplification of sentencing guidelines. Confusion matrices and summary statistics for both models are given below.

Custodial Sentence.
Full model
29 explanatory variables.

	Actual	
Predicted	Y	N
Y	413	81
N	45	98

Accuracy = 80%
Sensitivity= 55%
Specificity= 91%
Based on p=0.5 cut off.

Custodial Sentence.
Sparse model
19 explanatory variables.

	Actual	
Predicted	Y	N
Y	416	83
N	42	96

Accuracy = 80%
Sensitivity= 54%
Specificity= 91%
Based on p=0.5 cut off.

In summary, detecting or eliminating the existence of bias in the criminal justice system is very challenging but essential if the justice system is to be recognised as impartial. During this VSG we have investigated available datasets and identified approaches that could potentially help us understand the nature of causal links within the data.

Theme C

Sentencing is one of the most intrusive powers which impacts on both the offenders and the wider community. A few algorithms (AI and numerical models) are used in some countries; however, the result is not great. The implementation of AI is complex. Especially implementing AI needs to fit the need for accuracy and transparency to support penal legitimacy. Furthermore, the data needed to train any AI system is highly sensitive, and generally unavailable to an open group such as the VSG

The Challenges considered by this theme were:

1) Could the usage of AI reduce existing bias and arbitrariness in decision-making?
2) What is the extent of the bias in the judicial system?
3) What does consistency mean and how can we measure it?
4) What does proportionality mean and how can we measure it?
5) Can we use past judicial decisions to arrive at definitions of consistency and proportionality?
6) How might Statistical, AI or ML approaches be used to support consistency and proportionality in sentencing?
7) What are the other possibilities to use AI?

Note that both of the challenges 1 and 2 above linked to Theme B and there was a lot of cross discussion between the two themes on this point.

The limitations were: Data for training, selection of models, ethical and technical caveats.

1. **Pros and Cons analysis:**

The team drew up the following tables for the pros and cons of the various types of sentencing:

A) Pros and Cons of AI based sentencing

Pros	Cons
Requires almost no effort, fast,*which means less time to prepare evidence etc*	Different AIs might give different results, *can be good because of variability*
If well trained, they could potentially act following the law, *this can be harmful if laws are precedent-based, AI will follow itself*	They could give potentially bad sentences, *which might turn out to be good with time*
consistency side should be good, *again questionable if laws are even a tiny bit flawed*	Not adaptive to changing of society, *which menas less prone to populist trends*
impartiality, *lack of empathy*	standardisation of judgement, *more consistency*
	The output might be too vague, *allowing for changes and acts as a useful guidance*
	the same AI algorithm could give different outputs, *variability can be good for different reasons*

B) Pros and Cons of judicially based sentencing

Pros	Cons
They can act following the law and guidelines	Takes time and effort, cannot make accurate decsion when too many cases to judge
They can form a committee to make decisions on sentences	Biased sentences due to culture, political viewpoints
Empathy	Training background might vary
Traditional approaches, meaning many proper methodology had been developed	Debatable sentencing results. The sentencing was too strict or not.

C) Pros and Cons of Combining both

Pros	Cons
One gets different views on situations before making the final decision, *unless the entities start dominating or stop participating, for example trust all AI decisions*	Takes time before making sure that those algorithms are reliable, *can bring insights into new patterns*
Acting following the law is more likely potentially, *but can also impact abilities to change*	Judges could potentially ignore the AI simply because they feel they are more suitable for the task,*and the opposite is possible (doctors vs insurance in the US)*
AI algorithm could assist the process of analysing evidence and extract key data, *but the analysis it provides can be a black box*	How to validate the model is another question, *presence of a human can give a validity, which could be unjustified*

2. Quantifying sentencing process

Sentencing process is produced based on evidences and sentencing guidelines. To give a quantitative measure of the impact of the various types crimes in the given sentencing, a severity index model is proposed:

$S = a1\ a2*K + b1\ b2*M + r1\ r2*C + \ldots$

Where:

- K: number of people killed, a1 a parameter associated with the way killing was performed, a2 associated to the importance of crime with respect to other types of crimes.
- M: amount of money robbed, b1 how money is robbed, b2 the importance wrt other crimes.
- C: number of cars crashed, r1 severity of car accident, r2 importance wrt other crimes

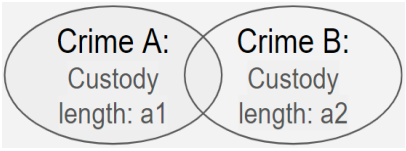
To elaborate this calculation, consider an example where there is a car accident. The sentencing of this would depend on many other details. For example, any casualty caused, whether any casualty are deliberate or by accident, assets value being damaged. Those considerations are combined linearly in this severity index. This is a way to quantify each big group of crime. However in many cases, there could be multiple crimes involved which needs a combined sentencing.

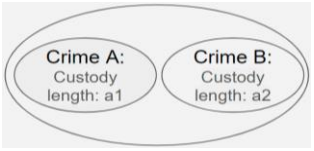
Linear combination of custody or not – combining two crime cases:

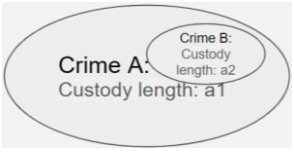
This leads to the question whether we should simply add different sentencing severities together. This is explained by those 3 Venn diagrams. Consider where a sentencing need to consider both crime A and B, and each along would be given an custody of length a1 and a2

respectively. However the final sentencing needs to follow complicated guidelines. In mathematical language, we can put these as whether A and B are joint or disjoint events.

When A intercept B is non-empty, then the custody length is w*a1 + (1-w)*a2 where w is a decimal between (0, 1). This is shown in diagram 1. When A intercept B is empty, then the sentencing length of custody is a1 + a2. This is illustrated the second diagram. If B belongs to A, then the total custody is only a1, which is explained in the third diagram.

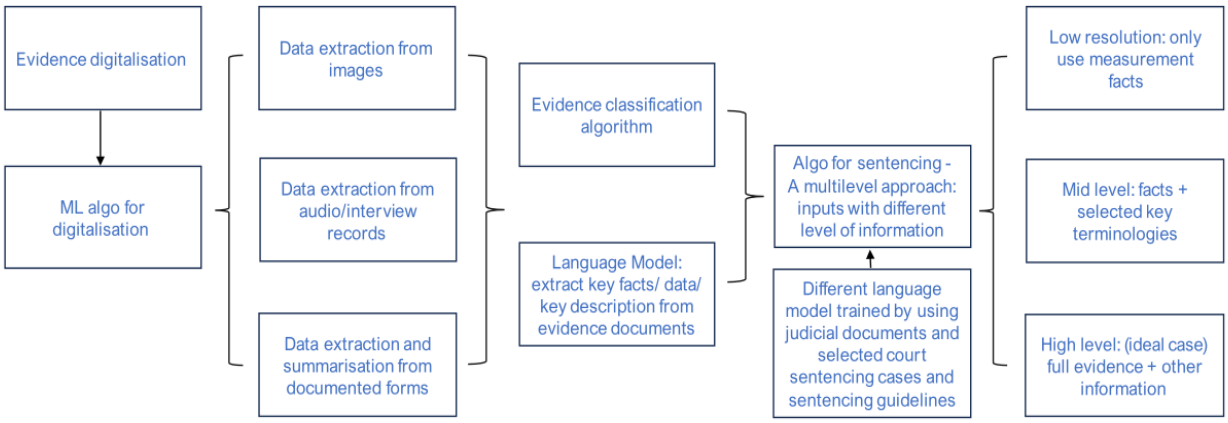






3.Hierarchy approaches and Suggested usages of AI in sentencing

Sentencing process is extremely time consuming. This is because this process is based on collecting, checking evidences, analysing evidences etc. Therefore, other than applying AI models directly on the final steps of sentencing, there are many steps worth considering to use AI models to speed up the entire process. This could assist to understand precisely where and how to use different AI models, we could consider Hierarchy approaches: Evidence filtering, evidence classification, metadata model. This flowchart gave a suggested pipeline where AI models could be used.



Other possible application of AI in assistance of sentencing include processing documents and cases.

5. Overall assessment

Overall the VSG achieved its primary purpose of bringing members of the academic and justice community together, identifying key areas of collaboration, giving academics access to interesting and informative data sets, and stimulating mathematical research in these areas. As can be seen from the summaries above, interesting results and questions in modelling, statistics and AI were obtained in each theme. As can be seen from the feedback below, the participants from the judicial system certainly found the meeting to be useful. The hope and expectation is that this meeting will stimulate future work, and collaboration in this important area.

6. Feedback from the participants

I look forward to some interesting follow up work.

- William

I would like to thank you for the opportunity to participate in this awesome event! I enjoyed it a lot.

- Julio

The UK is modernising the courts and tribunals system to make it more straightforward, accessible, and efficient. Understanding and optimizing the flow of cases and identifying potential bottlenecks is an important part in this ambitious programme. This workshop was a unique opportunity to tap into the rich experience from academics and industry professionals with a track record of applying mathematical approaches to tackle important challenges including the supply chain during the Covid pandemic or the optimization of industrial processes. The team concentrated on the Criminal Court System, first taking a Systems View breaking the flow of cases into its main stages and their connections. The team then used publicly available statistics to produce a high-level analysis of the system and implemented an agent-based model to simulate the flow under different limitations, such as the number of available judges or changes in the number of criminal cases. This could prepare important future work considering specific case characteristics using the large data sets available in the MoJ and HMCTS.

- Henryk

Work on sentencing disparities, which was carried out as part of 'Theme 2: Statistics' led to a rich discussion and a couple of concrete proposals. Professor William Lee developed a resampling approach with which to explore the robustness of the effect of

offender's gender on the probability of receiving a custodial sentence to potential unobserved confounders. If the mathematical underpinning of this approach could be formalised and its effectiveness therefore validated, I believe this could be a landmark solution with which to explore unwarranted disparities in sentencing and other decisions in the context of the criminal justice system and beyond. Another interesting application that stemmed from this theme was the model comparison undertaken by Dr Linda Nichols, who demonstrated that in predicting the probability of shoplifters receiving a custodial sentence, a model with 19 explanatory factors derived from the sentencing guidelines performed just as well as a model including 29 of such guidelines factors. This demonstration has a potential important application if conveyed to the right audience of criminal justice practitioners and Sentencing Councils and Commissions around the world. In particular, it seems that there is scope to simplify the sentencing process by highlighting the most relevant sentencing factors and potentially, removing from the guidelines other factors that are showing to have no effect on the final sentence.

- Jose

I have been working on matters relating to the use of AI in sentencing for a few years now and it was interesting to see how the issues were considered and tackled by the interdisciplinary group of mathematicians, statisticians and computer scientists. The team grasped the core legal issues straight away and quickly got to work analysing the pros and cons of various approaches and tackling the issues from some new and interesting angles. The potential solutions suggested were innovative and the scope covered surprised me, given the relatively short time-frame. The experience helped me reflect on the work that I'm doing, whilst also opening up opportunities to collaborate in the future. The way that the virtual workshop was set up worked really well, allowing for as much or as little engagement as people had time for. This presumably widened participation by allowing those balancing other responsibilities an opportunity to still take part and contribute. This seems an important point, given the benefits of engaging with a diversity of perspectives on these issues. I really enjoyed being a part of the virtual workshop and would not hesitate to participate again, if offered the opportunity.

- Lizzie

7. Follow up to the VSG

(i) The team from Theme A has met regularly since the VSG and is continuing to develop the ABM and dynamic models derived during the VSG.

(ii) A follow up to the VSG is planned. This will be a one day meeting to review the results obtained and to identify the next steps forward. This meeting will have an open invitation to both mathematicians and members of the justice system. The

venue and date of this is still to be decided, but we are in discussion with Lauren Hyndman to see if we can hold this at the ICMS. The remainder of the ICMS funds can then be used to facilitate this meeting.

(iii) The VSG has also generated a level of publicity. In particular, it was attended by the PLUS Maths journalist team. As a result two popular articles based on the theme of Maths for Justice, and in particular the use of AI in the justice system has recently appeared in the online mathematics magazine PLUS.

https://plus.maths.org/content/ai-be-judge-use-algorithms-criminal-justice-system

https://plus.maths.org/content/ai-be-judge-part-ii