The use of evidence generated by software in criminal proceedings

The Academy of Experts' Response to Call for Evidence

The Academy of Experts

This response has been prepared by an inter-disciplinary group of active expert witnesses drawn from the membership of The Academy of Experts. The Academy is a professional and accrediting body for expert witnesses of all disciplines. It is independently run by experts for experts and those using them.

The group includes IT practitioners and experts who have testified on criminal matters that rely on evidence derived from computer systems.

About The Academy of Experts (TAE):

- TAE is the premiere organisation for Expert Witnesses in the UK providing accreditation and training for Expert Witnesses and Mediators. Based in Gray's Inn, London, TAE was founded in 1987 with the objective of providing, for the first time, a professional body for experts to establish and promote high objective standards.
- All applicants to TAE who wish to become Accredited Practising Expert Witnesses undergo a rigorous vetting procedure to ensure standards of excellence are maintained. This is the process which gives the officially recognised full accreditation as a Practising Expert. Those achieving it are awarded designatory letters (eg SAAE, MAE or FAE). Ethical and professional standards are underlined by Codes of Practice and enforced by a disciplinary committee.
- Although there is representation on TAE's Council from the legal profession the majority of the officers, including the Chairman, are practising Experts - *The Academy of Experts is run by Experts for Experts and those using them*.

Response to Call for Evidence

It is the view of The Academy that the Presumption should remain, but it is vital to recognise its limitations and to welcome challenges to it. Issues of disclosure are central to addressing this situation.

1) The current common law (rebuttable) presumption is that computers producing evidence were operating correctly at the material time.

(a) Is this presumption fit for purpose in modern criminal prosecutions?

It is the view of The Academy that the presumption is generally fit for purpose in modern criminal prosecutions.

In coming to this view, we note the critical element of its being a *rebuttable presumption*. Its effectiveness depends on that being reasonably achievable.

(b) How easy or difficult do you believe it is at present for this presumption to be effectively rebutted?

There are currently barriers (identified below in response to 1) (c)) that interfere with the ability to rebut the presumption and thus make it more difficult than it should be. There are also specifics that applied in Horizon.

The first and most serious obstacle to rebuttal in Horizon was a failure of disclosure. This made it impossible to get under the covers of Horizon to question its reliability to an evidential standard. A suitable test system was not made available, and neither were error records until late and after criminal proceedings had concluded.

The second failure was the reliance by the Post Office on Gareth Jenkins of Fujitsu as an independent expert, when he was an employee of Fujitsu and designer of the system. He is being investigated for perjury with respect to the operation of the system.

(c) What barriers do you see in effectively rebutting this presumption?

Disclosure

There is a fundamental problem in the principle of disclosure of relevant material. There is a duty on the Prosecution to disclose any material that they think is 'relevant' to the case, which could potentially undermine the prosecution case or assist the case for the defence.

There is no such duty to disclose information about a computer system if the Prosecution does not consider such information to be relevant. The Prosecution is unlikely to consider information about a computer system to be 'relevant'.

However, experts instructed by the Defence are likely to disagree. They will consider the ability to question the reliability of the system and the associated presumption to be vital. The ability to do so effectively rests on:

- Appropriate disclosure including of vulnerabilities.
- Access to a reliable test system and associated test data to examine reliability. Reliability includes the degree of fidelity of replication of production conditions.
- Whether there have been changes over time or configuration that would affect results seen.
- Understanding of the process context of the system and its data.
- The full disclosure of known issues with the system that may lead to the validity of the system being questioned. Such evidence would include test reports and defects.
- Access to testing facilities equipped with the system in question needs actively to be provided by the courts where not volunteered by the parties.

We therefore propose that this (or a similar list) be included in a standard disclosure list for computer systems.

Most lawyers exercise their obligations to undertake disclosure in accordance with the rules with integrity. One of the issues we encounter is that an expert investigating the issues on which they are instructed will have a perspective and line of enquiry that the lawyer finds difficult to understand and thus to reflect well in what they disclose. An expert starts with a plan for investigation but adapts what is examined in the light of findings. Disclosing lawyers should take expert advice on the document and data types to be disclosed and disclose the whole system in its native form wherever possible, rather than documents extracted from systems. If the presumption is retained there needs to be a new presumption in favour of the defence being provided with access to the system in question.

If a party needs to get several court orders to force the other party to disclose then the judge should be asking far-reaching questions on why the documents were not disclosed in the first place. Was there an intention to withhold them?

(i) Please give examples where possible.

Within Horizon, there was disclosure of some flaws within the system and others were withheld. Information regarding what was known by whom and when was not disclosed. Such failures distorted the court's weighing what it was told and led to unsafe convictions. Amending the presumption would seemingly have little effect in such situations.

One of us reported in the case of transactions provided by a payments service provider for a defendant. The prosecution had disclosed little transaction data, adding more during the trial. Analysis of raw receipts that were provided by the defendants against the computer-reported transaction data showed that the columns of data had been "shuffled" so rendering them corrupt and meaningless. Each transaction line had the wrong combination of date, name and value. It was later discovered that the prosecution's expert had become aware of this fault but failed to mention it when giving oral evidence.

The elements in play here are the core software, the data within it, and the data produced. In this case, the output had been corrupted but the report was still produced by the prosecution as evidence.

Frustration of Investigation

Some defendants in criminal cases seek to frustrate investigation as a tactic in their defence. The burden required to overcome this barrier must be substantially reduced.

(i) Please give examples where possible.

There were egregious examples within the Post Office Horizon case in which access to Horizon test systems by the Postmasters' experts was prevented and the court failed to impose required access upon Fujitsu / the Post Office in a situation where the Post Office was acting as prosecutor.

Deliberate Corruption

Not all criminals tell the truth. There are occasions on which evidence is fabricated in an attempt to mislead the court. In such circumstances, the judge may support the jury in weighing what it is told to establish the facts. The "presumption" must not interfere with such open-minded consideration.

(i) Please give examples where possible.

The Horizon situation of Fujitsu's remote access to postmaster's accounts and manipulation of these is one example. The existence of the "back-door" was hidden.

A similar facility was available to one of us when working as an employee of a company that had implemented a retail point-of-sale system. They had the ability remotely to access and amend the branch's system without their knowledge.

One of us reported in the matter of disputed grain transport vessels in the Black Sea in the context of Russia being subject to sanctions. Ships carry Automatic Identification Systems (AIS) that transmit the location of the ship. On analysis, the data was found to have breaks and inconsistencies. Reports of location from anchorages did not consistently relate to the AIS data presented. It was concluded that although the computer producing the source data appeared to be operating correctly, the data had been manipulated later in its lifetime with the intention to deceive. Analysis required someone familiar with shipping operations to investigate the integrity of the AIS data to detect inconsistencies. The data had been obtained from an independent third party, downloaded and corrupted before being presented.

Test and the Validity of Results

Computer systems such as Horizon consist of multiple interacting elements. Computers are programmed by people: people are fallible.

In rare cases, systems are launched without any meaningful test having been conducted. These should be regarded as unreliable.

Proof of functional performance is routinely sought as part of systems development by conducting tests in which a range input conditions are considered, and performance is validated by comparing the output with the expected product.

Test involves the simulated reproduction of working conditions. Any such simulation will be an imperfect facsimile. In most conditions, the simulation is likely to be “good enough” to be useful. The range of conditions exercised can never fully replicate all that exist in the real world.

It is typically infeasible either to conceive or to test all the potential combinations of input conditions and transitory states (e.g. interrupted communication as in Horizon). The test result is therefore limited to an economic level of reassurance that the system is likely to perform effectively. Un-tested conditions will remain in which behaviour may be unexpected (e.g. the product of a calculation is inaccurate if the input is a negative number). Results from a well-conducted test are therefore usefully indicative but are never definitive. The courts must be warned of this limitation and of associated risks. In a large system with multiple components, it should not be presumed that a system should reliably perform under all circumstances. In the situation of a rebuttable presumption, ease of rebuttal becomes critical.

(i) Please give examples where possible.

One of us was asked to investigate a retail system. The retailer (R), who ran 3 or 4 convenience stores, claimed that the system was not collecting all of the sales data from remote locations.

The system supplier (S) claimed that they had over 100 installations and that none of the others had similar problems. They asserted that the problems must the result of external factors such as User Error, Power Supply fluctuations and telecoms failures.

The investigation did not find any evidence to support any of these explanations so the expert asked 'S' to arrange for a full system test during which one of their senior developers (their "Expert") would 'walk him through' the whole system. He also insisted that the test was done on the software that was actually installed in that user's site, despite 'S' saying that the software was standard across all user sites. The test revealed that the software on that one site had been modified by one of S's engineers in response to a request from the user when it was being installed, but the changes had never been tested in the development centre because they had never been recorded by the engineer. This resolved the question in dispute. This illustrates the need for joint testing.

Within Horizon, there was no definitive identification of cause, and this may remain unknown, with consequential limitation on ability to reproduce conditions giving rise to failure. In a complex system, such conditions may include rarely encountered combinations of inputs.

2) Are you able to provide examples from other jurisdictions or situations where the reliability of software must be certified?

No examples are submitted from other jurisdictions.

a) As examples of good practice?

Medical Devices in the UK must be certified to perform as expected before they can be deployed in clinical use.

Within computer forensics, the process by which data is gathered and measurements are taken must be assured. It is normal practice to require data to be processed by two different but accepted systems so that results may be verified.

When analysing a computer forensically, the first issue is obtaining a copy of the source system's data to analyse it. Is the copy a faithful facsimile? The second issue is that if laboratory computer systems are used to analyse the data, are they giving reliable results? In such analysis, the programmes used should be made available to both defence and prosecution, and results should be replicable and consistent between analytic systems.

Within the science of metrology, machines used for measurement are required periodically to be calibrated against an established standard. The results of measurement are relied upon only within specified working conditions. Operating within such conditions is regarded as demonstrating that a device is reliable. Corroborating evidence is required to show that the claimed conditions actually applied.

When considering the reliability of a software programme, one can assess its development and testing history, the process by which it was brought into being. Software within the nuclear engineering industry is subject to rigorous process checks to assure its reliability and quality.

Many issues with computer systems arise from failures in design. These do not give rise to "defects" as the system may still perform to specification, but the design does not reflect the intention of users.

b) As examples of things to be aware of?

Software development uses "versions" or "revisions" of software to record changes, such as the application of bug-fixes. A "configuration database" is sometimes used to record changes to revisions, when and where they are applied. Maintaining the integrity of configuration data can be a complex undertaking that is often done unreliably. In the absence of such data, one cannot be sure that attempts to recreate the situation that applied at a time of interest will be reliable.

Where cloud software is used, as well as some on-premises software, the rate of update is high and there can be uncertainty as to the configuration of the system at the relevant time. Updates are forever being made both to add functionality and to fix bugs. There is thus great operational difficulty in reproducing the system in a condition that reasonably faithfully represents the condition of the time.

3) If the position were to be amended, what in your opinion would be the most appropriate and practicable solution given our aims and objectives set out above? It would be helpful if your answer could address as many of the below as possible:

a) What procedural safeguards need to be in place to ensure your proposed solution is effective?

Multiple as in previous responses – test, configuration, calibration, process validation.

b) How might we ensure that any proposed solution is, as far as is reasonable possible, future-proofed?

Traditional software is deterministic – specified inputs result in the same output when tests are repeated. Artificial Intelligence and some associated methods are probabilistic – specified inputs may result in a distribution of outputs when tests are repeated. The nature of the system needs to be considered before interpreting the test results.

c) How might we ensure that any proposed solution is operationally practical?

Some differences between the test system and the live production matter, others do not. This is an issue of expert judgement in the context of the matter.

An organisation may know what the system should have done. There are issues in computers not always keeping complete records of the text of messages that were actually sent out. An organisation may seek to rely on what the system does today as evidence of what it did years ago. This is unreliable where there has been change.

d) If your proposed solution requires the use of expert witnesses (either jointly or singly instructed), what expertise and qualifications would that person require? To your knowledge are there sufficient such people at present?

An expert witness needs to have been trained and assessed in practice as an expert witness. Recognised expert witness bodies such as The Academy of Experts undertake this.

Those designing the tests need to be familiar with the systems being examined, the business and operational processes embedded within the systems, and with good practice in IT testing, preferably gained from practical experience and knowledge of accepted practice. Appropriate professional qualification would include Chartered IT Professional status awarded by BCS and certification under ISTQB.

There are experienced expert witnesses practicing in the UK. There is an issue in attracting sufficient to publicly funded criminal work as the rates paid are significantly below what many may earn elsewhere.

Understanding of Process

When it comes to putting appropriate interpretation on data related to a particular area of expertise, it is important that the expert understands the process within which the data was collected. For accounting data, the expert should be, for example, a chartered member of a recognised accounting body, or someone who has appropriate experience and/or holds qualifications in a relevant quantitative financial discipline, who can thus pass opinion on the accounting context of the data concerned. Appropriate interpretation can only be assured in the context of a known process that establishes and preserves the provenance of the information.

Example of Process

One of us investigated a case of suspected fraudulent refund in retail. Under cross examination the retailer's auditor admitted that the store manager, deputy and supervisor (who did the end-of-day validations) were not checking that every refund was signed by both the person who did it and countersigned by one of them.

The prosecution relied on a report that showed all of the transactions were done with the same Staff ID and a correct passcode. But every receipt showed the staff ID, so any member of staff could use anyone else's staff ID if they knew their passcode. The expert demonstrated a simulation of this compromise to the jury, so showing that the reliance on Staff ID and Passcode to be insufficient.

A second case concerned alleged false charges by a bank in payment card processing. In the course of using a card for payment, a counter within the card is incremented and the value of the counter is stored in relation to the transaction. Knowledge of this process was used to show that an individual's card had been used by another member of the cardholder's family without their knowledge, incurring transactions that were disputed. On disclosure of this information, the cardholder's claim for refund from the bank was withdrawn.

This is another example of a situation in which the corroborating conditions that were claimed by one party to have applied were shown by the other party to have been absent.

A case of alleged fraud in logistics came down to the units that were being counted. Just as half-a-dozen eggs within an egg-box can be counted as six units, or as one box. Interpretation of the data is vital to making sense of the data reported.

4) In your opinion, how should 'computer evidence' for these purposes be best defined?

This is far from straightforward. The definition proposed for evaluation is noted.

Regarding the validity of the presumption, bespoke programs are common and often generated for specific purposes, at least in some industries. They range from high level, including research, to quite simple spreadsheets. The validity of the presumption to such systems is highly questionable. Their provenance must be stated; they must be validated and checked (commonly against recognised standards) and open to reasonable interrogation in case of reasonable doubt, so any presumption is not abused. The full input and output must be available if reasonably requested. Mistakes, even by quite senior staff, are not unknown. Spreadsheets should normally be in source (xls) form to allow review – not pdfs.

a) Do you agree that evidence generated by software, as set out above, should be in scope, and that evidence which is merely captured / recorded by a device should be out of scope? Please provide a rationale for your answer.

Any such definition is likely to be problematic and may be avoided if the presumption remains. As even data capture devices commonly contain software, defining them in a way that distinguishes from accounting systems is difficult. Any change would be a prompt for re-calibration and validation. This is likely to be a problem in the context of frequent update (perhaps several times a day).

There remain some instruments (e.g. a mercury thermometer) that are not digital, but they are few. A stand-alone instrument may contain a computer and software and still perform the function of stand-alone and dedicated measurement, such as a speed-camera.

The text refers to “software”. The performance of systems is affected by hardware on which the software operates. It is therefore recommended to refer to systems that incorporate multiple components, hardware, communications networks, data and software, and to give examples.

Routine and ‘industry standard’ computer programs are normally validated by use where widely deployed. The presumption is then reasonable, subject to rebuttal where relevant to the issues and valid ‘reason to suppose’.

i) Can you provide specific examples of the type of evidence you believe should be in scope?

Accounting data from corporate accounting and financial management systems.

ii) Can you provide specific examples of the type of evidence you believe should be out of scope?

The exclusion of data captured from devices such as breathalyser and the other examples given in the consultation is reasonable and practicable in avoiding injustice. Definitional issues are likely to be rife.

5) Are there any other factors which you believe are important for us to consider?

One of the serious failings in relation to Horizon was the lack of independence of Fujitsu’s Gareth Jenkins when acting as an Expert in the original criminal prosecutions of Postmasters. The subject of expert independence is a well-known one, the general line being that it affects weight but not admissibility. In situations where the evidence is not controverted, there are risks that the weight is not assessed.

Our view is that there should be a presumption for expert evidence being provided by both sides and that courts should take steps to ensure that all available expert evidence is tested in court. The strategy of advocates in managing their presentation should not be allowed to interfere with this (e.g. Lucy Letby case).

We perceive that there are risks in regarding Horizon as a one-off with unique circumstances, that are assumed not to recur. We believe that this would be a mistake and that systematic issues of non-disclosure, reliance on an insufficiently independent person to provide expert opinion and insufficient access for testing are likely to be seen elsewhere, as they have in some of the examples cited above.