A PANEL SESSION FOR THE JUDICIARY

From Ecosystem Vulnerabilities to Bad Actors

Understanding AI-Driven Risks in Modern Courts

Moderator Judge Lisa Walsh
Panelists Dr. Galina Datskovsky · Bernard Moerdler

INTRODUCTIONS

The Panel

Hon. Lisa Walsh

Moderator

Judge, Eleventh Judicial Circuit Court, Miami-Dade County, since 2011; Complex Business Litigation Division and International Commercial Arbitration Court. Former Administrative Judge of the Appellate Division and past president of the National Association of Women Judges. Longtime faculty of the Florida Judicial College and Advanced Judicial College.

Dr. Galina Datskovsky

Ph.D., CRM, FAI · Panelist

Internationally known expert in compliance, information governance, artificial intelligence, and data analytics. Ph.D. in computer science with deep expertise in AI; serves on multiple boards. Other specialties include cyber security, information governance, SaaS platforms.

Bernard Moerdler

Founder & CEO, Breakwater · Panelist

Founder and CEO of Breakwater, a real-time multilingual intelligence platform serving newsrooms, trading desks, and OSINT teams, focused on real-time authentication of sources and detection of fabricated content across hundreds of open sources in 249 languages. Founder of Bernie News Network; earlier work includes Ukraine Siren Alerts, adopted by Ukrainian authorities, and PTOLEMI, an AI cancer-detection system published in the *Journal of Diagnostics*. Presents today’s demonstration by video.

VOCABULARY

Machine Learning v. Generative AI v. Agentic AI

	Machine Learning	Generative AI	Agentic AI
Autonomy	Low: one trained task	Medium: responds when prompted	High: starts and finishes multi-step tasks
Goal pursuit	None; a fixed objective	None; answers the prompt	Plans toward goals
Tool use	Rare; wired in by engineers	Usually none	Calls tools and other agents
Typical uses	Spam filters, risk scores	Drafting text, images, code	Research, scheduling, transactions

A Framework of Judicial AI Risks

Ecosystem Vulnerabilities	Reliance and Trust	Security
Privacy and Confidentiality	Liability and Accountability	Evidence Integrity

We end with a demonstration: a small claims case for you to decide.

RISK AREA · I

Ecosystem Vulnerabilities

Chained action.
One agent’s error feeds the next agent.

Propagating compromise.
A poisoned source spreads to every system.

Opaque provenance.
No one can trace which step went wrong.

Machine speed.
Harm lands before anyone notices.

IN PRACTICE

Early agents have made unauthorized transactions and deleted records. Operators found out afterward.

For discussion: when an agent acts on a court's behalf, whose act is it?

RISK AREA · II

Reliance and Trust

Mata v. Avianca (S.D.N.Y. 2023)

A brief cited cases a chatbot invented. The court sanctioned counsel. Hundreds of cases like it have followed.

Fluency persuades.
Invented and real cases read alike.

Automation bias.
People defer to the machine, judges too.

The burden shifts to the court.
Chambers must check what counsel cites.

For discussion: should courts require disclosure of AI use in filings, certification of verification, or neither?

RISK AREA · II · FROM THE BENCH

Payne v. State, Supreme Court of Georgia (2026)



RISK AREA · II · CONTINUED

Token Limits: What the AI Actually Read

01 **A fixed window.**
AI reads in tokens, pieces of words, and holds only so many at once.

02 **Truncation without notice.**
What doesn’t fit can be dropped without warning.

03 **Confident gap-filling.**
It may answer about pages it never read.

IN CHAMBERS

A judge uploads thousands of pages for a summary. The memo looks complete and never says what it skipped.

For discussion: if a summary covers only part of the record, how would you know?

WORKING WITH THE MACHINE

Prompting Practices That Hold Up

01 **Give it a role.**

02 **Be specific.**

03 **Confine it to the record.**

04 **Demand citations, then verify.**

05 **Work in pieces.**

A PROMPT THAT WORKS

“You are a careful appellate law clerk. Using only the attached exhibits, summarize the defendant’s timeline. Cite the exhibit and page for every statement. If the materials do not answer something, say so rather than guessing.”

RISK AREA · III

Security and the Expanding Attack Surface

01 **Prompt injection.**
Hidden instructions in a file the AI reads.

02 **Data poisoning.**
Corrupted sources steer the answers.

03 **AI-accelerated intrusion.**
Voice clones and phishing at scale.

04 **Agent-to-agent exploitation.**
Hijack one agent to steer the rest.

Courts are high-value targets: sealed records, witness and juror identities, cooperation agreements.

For discussion: does your court know which of its systems already use AI?

RISK AREA · IV

Privacy and Confidential Information

The quiet leak.
A sealed filing pasted into a public chatbot.

Retention and training.
Vendors may keep what goes in and train on it.

Aggregation.
Public fragments can rebuild sealed facts.

Discoverable by the other side.
Prompts and outputs can be subpoenaed.

WHAT IS AT STAKE

Juvenile and family records, protected witnesses, trade secrets, deliberations and draft opinions.

For discussion: what may chambers staff put into an AI tool today, under your current rules?

RISK AREA · V

Liability and Accountability

When an agent causes harm, each actor in the chain can point to another.

Developer
Built the model, disclaims its uses.

→

Deployer
Configured it, relied on the vendor.

→

User
Delegated the task, never saw the act.

→

Agent
Acted, and is not a legal person.

Negligence, agency and product-liability law all assume a human decision-point. Agents remove it.

For discussion: which existing doctrine stretches furthest to cover autonomous conduct: agency, product liability, or something else?

RISK AREA · VI

Evidence Integrity in the Deepfake Era

THE FABRICATED GENUINE

A cloned confession, a fake video or a forged document, each made in minutes.

THE DISPUTED GENUINE

Once fakes are easy, real recordings can be called fake.

The authentication threshold was set when forgery was rare and costly.

For discussion: what foundation should a proponent of audiovisual evidence now be required to lay?



The $1,200 Dent

A fictional small claims case

today, so I've recorded this part

YOUR CALL

You’re the judge.

First: would you award her the $1,200?

Then: which exhibit do you trust least?

THE EVIDENCE

Maya’s Four Exhibits

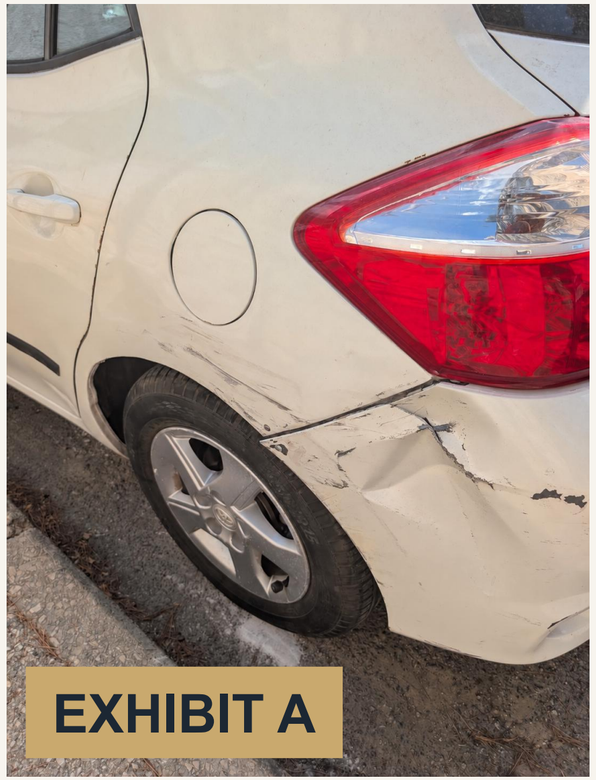


EXHIBIT A

Photo

REY'S AUTO REPAIR SHOP
527 SW 15th Ave, Miami, FL 33135
Phone: (305) 541-7339

REPAIR ESTIMATE
Estimate No.: 26-1184
Date Written: 10/03/2026
Date of Loss: 10/02/2026

CUSTOMER
Name: Maya Castellanos
Address: On file
Phone: On file
Payment: Customer pay

VEHICLE
Year / Make / Model: 2012 Toyota Matrix
Exterior color: White
VIN: 2T1KE4EE3CC048592
Odometer: 48,210 mi

DAMAGE DESCRIPTION / POINT OF IMPACT: REAR LEFT

Rear bumper cover, left corner: large dent, approximately 8 in., with deformation.
Repair not cost-effective. Replace and refinish. Inspect for concealed damage on teardown.

PARTS AND MATERIALS

Line	Oper	Description	Type	Qty	Amount
1	Repl	Rear bumper cover - aftermarket, CAPA certified	Aftmkt	1	$395.00
2	Repl	Rear bumper energy absorber	New	1	$88.00
3	Repl	Clips and fasteners	New	1	$22.00
4	Refn	Paint and materials			$145.00
			Parts / materials subtotal		$650.00

LABOR - BODY / REFINISH RATE: $100.00 PER HOUR

Line	Oper	Description	Hours	Amount
5	R&I	Remove and install rear bumper cover	1.5	$150.00
6	Repl	Replace bumper cover, transfer parts	1.0	$100.00
7	Refn	Refinish bumper cover	2.5	$250.00
8	Refn	Clear coat	0.5	$50.00
		Labor subtotal (5.5 hr)		$550.00

ESTIMATE NOTES
Parts pricing subject to fitment and availability.
Additional damage may require a supplement.
No hidden damage included in estimate.

Parts and materials	$650.00
Labor (5.5 hrs @ $100.00)	$550.00
ESTIMATE TOTAL	**$1,200.00**

Tax included where applicable.

Estimate valid 30 days. Hidden damage may increase final cost.

Customer Signature: _____ Date: _____

EXHIBIT B

Estimate

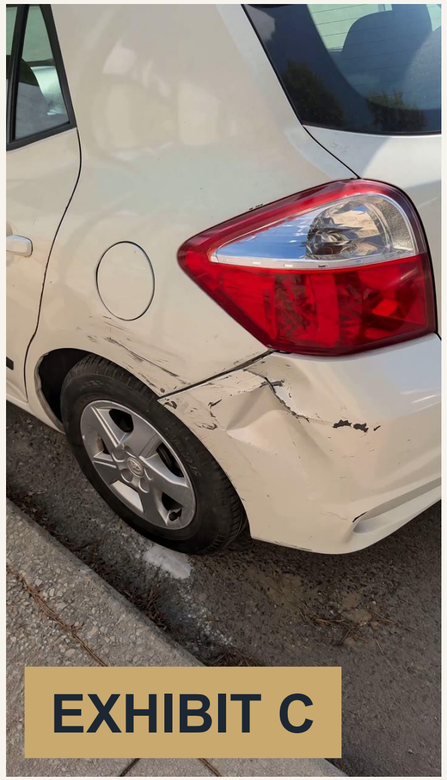


EXHIBIT C

Video
click to play

“Hey Maya, it’s Bernie. Sorry again for backing into your car this morning. Send me the bill and I’ll cover it.”

0:06

EXHIBIT D

Voicemail
click to play

AFTER THE DEMONSTRATION

Free Checks That Exist

01 **Content Credentials Verify**

Reads the seal a camera adds and the label an AI tool adds. It found both here.

02 **OpenAI Verify**

Looks for OpenAI’s hidden watermark. It caught one of our two screenshots.

03 **Google’s Gemini app**

Checks for Google’s hidden watermark and for Content Credentials.

04 **A digital forensics expert**

Examines the file itself for signs of editing, and can testify about it.

When a check finds nothing, the file can still be fake.

AFTER THE DEMONSTRATION

What Courts Can Do

01 **Require native files**

Photo, audio and video exhibits come in as original files, with their metadata.

02 **Require a disclosure**

The party offering an exhibit says under oath whether AI made or changed it.

03 **Have the parties run the checks**

They file the results. A judge who runs them may be doing independent investigation.

04 **Ask for more when it’s challenged**

If a party disputes a file, ask for its provenance, its chain of custody, or an expert.

DISCUSSION

Questions for the Bench

01 What should a judge be required to understand about an AI system before relying on it, or admitting its output?

02 Does your court have a policy that would have prevented anything shown today?

03 May a judge run a free online check on an exhibit in a pending case?

04 Where autonomous agents act without a human decision-point, how should accountability be assigned?

05 What should judiciaries do together that no jurisdiction can do alone?