WONK PREMIUM BRIEF:

Balancing The Promise Of AI With The High Stakes Of Child Welfare

What Happens When Agencies Adopt Generative Ai Before They Understand It?

BY SARAH CATHERINE WILLIAMS

Child welfare agencies across the country are confronting the same reality: too few staff, too much paperwork, and decisions that carry life-altering consequences for children and families.

Caseworkers spend countless hours documenting visits, preparing court reports, navigating policy manuals, and updating systems—time often pulled away from direct work with families.

In that environment, the appeal of generative artificial intelligence (AI), like ChatGPT or Co-Pilot, is obvious.

But in child welfare, efficiency is never the *only* question, and its purpose is in service of the direct work with families.

When leaders plan their approach for adopting AI in practice, they can avoid having urgency dictate strategy by putting AI in service of their goals, rather than have it feel like a trend happening *to* them.

When agencies adopt technologies they do not fully understand—or deploy them without clear guardrails—they risk importing new forms of error, bias, and harm into one of government's highest-stakes systems.

THE APPEAL OF THESE TOOLS IS REAL

The interest in generative AI is not about replacing caseworkers. It is about helping them manage workloads that have become unmanageable.

Caseworkers spend enormous amounts of time documenting case histories, preparing court materials, consulting policy and practice manuals, and entering information into case management systems.

Every hour spent on paperwork is an hour not spent with children and families.

Used thoughtfully, AI could help reclaim some of that time. Agencies are already exploring—and tech companies are pushing—tools that can, for example,convert handwritten notes into digital records, generate referrals based on case information, or assist with case planning.

The strongest case for AI in child welfare is simple: if technology can absorb routine administrative tasks, workers may have more capacity for the human-centered and relational work only people can do.

THE RISK IS REAL, TOO

Every promising use of generative AI is matched by a corresponding risk.

Generative AI tools are emerging technologies. They evolve rapidly, often without independent evidence demonstrating reliability in child welfare settings.

Many are produced by private companies using proprietary systems that public agencies cannot fully inspect or control.

They are also prone to hallucinations—outputs that sound legitimate but are false.

A fabricated citation in an academic setting is embarrassing. A fabricated fact in child welfare could shape an investigation, a removal decision, and a family's future.

Amplification of pre-existing bias is a well-established concern in policy arguments around other algorithmic approaches like predictive analytics.

Generative AI has its own distinct concerns, based on whether the model's underlying training data reflect biases that could distort how it evaluates, analyzes, and presents data.

In a field where mistakes can be irreversible, experimentation without safeguards carries real costs.

WHY AI LITERACY MATTERS IN CHILD WELFARE

Many workers may reasonably assume that if they ask an AI tool to summarize case notes, the result will be accurate, complete, and factual.

But that is not how these tools work. They do not verify truth as a human would. They generate language based on learned statistical patterns.

That distinction matters. A summary may appear polished while omitting a critical detail, overstating a concern, or inventing information altogether.

A tool could capture 95 percent of a family's history correctly and fabricate the remaining 5 percent. In child welfare, that remaining 5 percent may matter most.

Under time pressure, workers may not catch subtle errors. The result can be documentation that misrepresents what happened in a family's case—with legal, safety, and ethical consequences.

This is why AI literacy matters. Staff need to understand not only how to craft prompts, but also how to verify outputs and recognize their limitations.

WHAT HAPPENS TO PROFESSIONAL JUDGEMENT?

There is also a workforce development question that receives less attention.

If staff come to rely heavily on AI-generated summaries, plans, or recommendations, what skills could weaken over time or never be developed in the first place?

Professional judgment in child welfare is built through experience and supervised reflection. It requires learning how to synthesize messy information, weigh risk, and understand nuance.

When staff learn to think with large volumes of underlying information, they learn to think about decisions based on that information.

While that may not be the only way to learn that decision-making process, it does mean that cultivating judgment becomes even more important.

Outsourcing synthesis and analysis to AI tools can help make good decisions, but outsourcing decisions is a line agencies should not cross.

PRIVACY AND CONFIDENTIALITY IS NOT A SIDE ISSUE

Protecting confidentiality and privacy is a foundational issue in the use of generative AI in child welfare.

Yet many of the most accessible AI tools are commercial products built by private companies using proprietary technology.

If workers feed names or other identifying information into unsecured public tools, agencies may lose control over how that information is stored, processed, or reused.

That makes training and procurement protocols for proprietary public systems essential.

Most child welfare staff have never been formally trained how to use AI responsibly, much less the data security or privacy risks associated with these tools.

Child welfare researchers have used local large language models operated on infrastructure they control to protect data and personally identifiable information.

But for many jurisdictions who would like to use that approach, building and maintaining those systems may be financially or technically unrealistic, and require novel policy approaches.

If agencies cannot protect family information, they are not ready to deploy these tools.

WHEN TECHNICAL ERROR BECOMES HUMAN HARM

In child welfare, the stakes when an error occurs are unusually high and the harm may be irreversible.

A child could remain in danger, or family could be unnecessarily separated. A court could rely on flawed information.

Unlike some predictive analytics tools that produce visible scores or rankings, generative AI errors can be harder to detect.

There may be no obvious marker separating accurate content from fabricated content. A polished narrative can conceal falsehoods better than a flawed spreadsheet ever could.

Reviewing outputs carefully requires experience and time—one of the very resources these tools are meant to save.

That creates a structural tension agencies must confront honestly, and raises the stakes for judgment as a core function of the workforce.

The opportunity now is to think proactively about governance *before* serious harm occurs, rather than after.

ACCOUNTABILITY CANNOT BE RETROACTIVE

No single person is solely responsible for AI-generated harm. Accountability can span frontline workers, supervisors, agency leadership, and vendors.

That is why governance structures are important. Agencies need rules in place before deployment that clarify:

- Which tools may be used and for what purposes
- When human review is required
- Who is responsible for monitoring tool performance
- What families are told about AI use
- What evidence of harm triggers investigation
- What remedies are available when harm occurs

Without those answers, accountability becomes almost impossible after the fact.

THE REAL DECISIONS BEFORE LEADERS

Before implementing AI-assisted tools leaders will want to contemplate and plan around several core questions.

Problem Definition: What problem are we solving, is AI the right tool, and how do we measure success?

Implemented responsibly, in consultation with frontline workers and evaluators. agencies can solve the efficiency problem.

Agencies that cannot name a specific problem and how AI will solve it risk investing limited resources in solutions in search of a problem.

Tool Design and Validity: What data trained this system? Does it reflect our population? What outputs exactly is it producing, and how reliable are they?

Independent validation –from someone other than the vendor–is vital for agencies to understand where and for whom the tool fails.

Otherwise, AI failures remain likely invisible until a family is harmed.

Ethics and Civil Rights: Who benefits when AI functions as intended, and who bears the risk when it fails? What protections ensure that risk doesn't spread unevenly based on race, disability, income, and geography?

This is a governance question that can shape these tools toward not just preventing amplification of existing biases and disparities, but reducing them.

Getting it right will require outside expertise, including from impacted families and communities and experts in civil rights and due process.

Practice and Workflow: Does the tool strengthen or weaken professional judgment? Does it reduce time with families or increase it? What input do frontline workers and supervisors have in implementation design?

Promoting AI literacy in the child welfare workforce can preserve professional judgement and mitigate the over-trust workers have in algorithmic outputs.

Privacy, Data, and Security: What information do the tools collect, where does it go, and does its storage and use comply with confidentiality law and policy?

Experts in data privacy and security can provide critical input needed to protect sensitive information about families.

Governance, Accountability, and Oversight: How do you monitor use and applications, what disclosures will you make to families, courts, and legislatures, and who holds accountability?

Accountability that diffuses across workers, vendors, and courts never fully lands anywhere, reducing responsiveness to a critical responsibility. Governance protocols make accountability clear and legible outside of adversarial processes.

WHAT ACTUALLY CHANGES NOW

The debate over AI in child welfare is often framed as innovation versus resistance. That framing misses a more subtle–yet nonetheless vital–nuance.

The real issue is whether agencies can use new tools without compromising fairness, privacy, judgment, or trust.

Generative AI may eventually help relieve administrative burdens and improve practice. Its integration into practice isn't in service of mere efficiency, but an expanded focus on deeply human work.

Empathy, moral responsibility, and relationships are at the center of child welfare work, elements no tool can replace.

About Child Welfare Wonk

Child Welfare Wonk is a strategic policy intelligence platform serving a community of over 3,000 decision makers across child and family policy.

Our analysis informs the work of policymakers, providers, philanthropies, and advocates across the national, state, and local levels.

We offer the intelligence layer decision-makers need; analysis and projections grounded in historical context, real-time insight, and future-focused strategy.

We provide a free weekly newsletter and podcast, member-only premium resources and events, and organizational partnerships that integrate our intelligence and sense-making into your strategy.

To learn more about and join our premium community, visit the *Wonk Briefing Room*.