SunCloud Health

YOUR VOICE. YOUR CHOICE. OUR TEAM.

SunCloud Health Continuing Education Series

Transdiagnostic, integrated care for people with complex co-occurring disorders.

Centering Survivor Agency:

Ethical integration of AI in therapeutic and violence-prevention practice

Aislinn Conrad, PhD · University of Iowa School of Social Work

SUNCLOUD CME · NOVEMBER 18, 2026

Centering Survivor Agency

Ethical integration of AI in therapeutic and violence-prevention practice

Aislinn Conrad, PhD · University of Iowa School of Social Work

BY THE END OF THIS SESSION

Three learning objectives

1. Identify the primary ethical risks and benefits of incorporating AI-based tools into therapeutic and violence-prevention work.

2. Describe practical strategies for preserving client and survivor agency, consent, and voice when AI is introduced into clinical decision-making.

3. Apply an evaluation framework to assess whether a given AI tool is appropriate for vulnerable or trauma-exposed populations before adopting it into practice.

ROOM POLL · ANONYMOUS

Who is in the room?

Your role

Clinician · Advocate · Supervisor · Case manager · Student · Other

Your setting

Violence prevention or shelter · Counseling or therapy · Child welfare · Health care · Other

Your comfort with AI

Personal use, 1 to 5 · Professional use, 1 to 5

Results are shared only in aggregate. Demographic questions are optional.

BACKGROUND

What do we mean by AI?

Broad

Any computer system that performs tasks usually requiring human judgment, such as sorting, predicting, or recommending

Session

Tools that learn patterns from past data and use them to score, sort, flag, or generate something new

Practice

1. Risk scores and screening tools
2. Chatbots and client-facing assistants
3. Summarization and documentation tools

Together, they reflect the data they were trained on, and the people who chose that data and its purpose.

ROOM POLL · 1 OF 2 · ANONYMOUS

Is AI being used by your employer right now?

- Yes
- No
- Not sure

"AI" includes scheduling tools, documentation assistants, risk scores, and chatbots.

ROOM POLL · 2 OF 2 · ANONYMOUS

How is AI used at your agency?

Choose one
Not used
Planning or piloting
Documentation, scheduling, or summaries
Risk screening or decision support
Client-facing tools, such as chatbots
Don't know

CHAT · ABOUT 3 MINUTES

Why does this topic matter to you?

Type one sentence in chat

Why does this topic matter for you, your clients, or your agency?

No wrong answers

We will read a few responses aloud later, without names unless you choose to share one.

WHERE AI IS ALREADY USED

AI is already in violence-prevention work

Tool	Setting	What it does
Allegheny Family Screening Tool	Child welfare hotline, US	Predictive risk model scores maltreatment referrals
Lizzy	Counseling centers, Germany	AI risk analysis when victims seek help; required by law
Hope Chat	Survivor support, US	24/7 chatbot connecting survivors to resources
Non-AI Comparison		
Dash	Police and social services, UK	Checklist; Minister for Safeguarding has called for an overhaul
EPV risk tool	Intimate partner violence, Basque Country	Algorithm scores severity from 20 questions, and judges see the result. Critics question the subjective items and the evidence behind it.

Sources: Chouldechova et al. (2018); Conrad (2026); deutschland.de (2026); Valdivia et al. (2025); Alvarez Sanchez; The Week (2026); Domestic Shelters webinar listing.

CASE · MAYA, 30 · HANDOUT

Her situation

Maya is 30 and lives in a rural county with her two-year-old son. She spent part of her own childhood in the child welfare system. She has a bipolar diagnosis and sees a psychiatrist in town. Her son's father has been abusive, and their relationship has been on and off. She has left him again and is staying in a shelter with her son. She has no formal education and minimal work experience. The public counseling center where she has come for help is in the same building as her psychiatrist.

The tool (fictional)

Name: Intake Safety Score

Purpose: estimates the chance of serious harm in the next six months, as a score from 0 to 100 with a band: low, moderate, or high. Used at intake to set priority for shelter placement, legal advocacy, and multidisciplinary case review.

Inputs: prior child welfare involvement, public benefit receipt, housing instability, police calls at the address, mental health diagnosis from records, and the partner's prior charges.

Design: built by a private vendor on regional administrative records.

Testing: pilot with about 40 survivors at one urban program. Not peer reviewed. No results by subgroup.

Bias audit: none published.

Maya's score: 62, moderate.

Composite fictional case. Not based on any real agency or product.

LEARNING OBJECTIVE 1

Identify the primary ethical risks and benefits of AI

Where the risk comes from, and what AI can offer

OBJECTIVE 1 · ETHICAL RISKS

Four ways structure becomes harm

1 · Carceral logic

A risk label authorizes more monitoring and reporting, without more support.

2 · Structural bias

Contact with systems, such as benefits or prior reports, becomes a risk signal.

3 · Epistemic injustice

A survivor's account is discounted relative to an opaque score.

4 · Psychological harm

Constant monitoring echoes coercive control and suppresses help-seeking.

Conrad (2026), Journal of the Society for Social Work and Research, 17(3).

HARM 1 OF 4 · CARCERAL LOGIC

A high-risk label can mean more monitoring, not more help

What it is

Once labeled "high risk," families, partners, or caregivers may face more monitoring, documentation, and reporting, without a corresponding increase in material support.

In Maya's case

Her band of 62 could add documentation and reporting, while her requests for childcare and housing wait. The report would reach a mother who was herself in the child welfare system.

In the field

The Allegheny Family Screening Tool scores child maltreatment hotline referrals. A higher score can change how a referral is handled (Chouldechova et al., 2018).

Restated threshold: AI must not independently trigger investigation or surveillance without human deliberation.

Conrad (2026); Benjamin (2019); Brayne (2021); Chouldechova et al. (2018).

HARM 2 OF 4 · STRUCTURAL BIAS

Poverty and system contact become risk signals

What it is

Administrative data record contact with systems. Poverty-linked contact, such as cash assistance, housing instability, or prior reports, is then read as individual risk. Black, Indigenous, and low-income families as most affected.

In Maya's case

Her childhood in care, her benefit receipt, and her shelter stay can each add points. None of them shows what her partner did.

In the field

A health algorithm used past spending as a stand-in for need, and systematically underestimated the needs of Black patients

High overall accuracy can coexist with significant disparate impact

Conrad (2026); Obermeyer et al. (2019); Chouldechova et al. (2018); Dettlaff et al., 2020; Obermeyer et al., 2019

HARM 3 OF 4 · EPISTEMIC INJUSTICE

The score carries more weight than the survivor's own account

What it is

Complex survivor narratives are translated into standard categories such as risk, compliance, or threat. Machine outputs carry more institutional authority than survivor testimony, so the survivor is discounted as a source of knowledge (Fricker, 2007; Dotson, 2011).

In Maya's case

Her account of why she left and what she fears is coded into categories. When the score and her account disagree, the score wins by default.

In the field

The UK's Dash checklist is not AI. Practitioners are still encouraged to override low scores with professional judgment. A UK minister has said it "doesn't work." Authority does not guarantee accuracy.

Restated threshold: survivor voice must be able to change the outcome, not only be recorded.

Conrad (2026); Fricker (2007); Dotson (2011); The Week (2026).

HARM 4 OF 4 · PSYCHOLOGICAL HARM

Being watched can feel like the abuse itself

What it is

Trauma involves loss of control, unpredictability, and involuntary monitoring. Opaque alerts and continuous assessment can reactivate those responses. Perceived monitoring alone can discourage disclosure and suppress help-seeking.

In Maya's case

Maya has a bipolar diagnosis and a history of system involvement. Knowing a score follows her could lead her to stop telling her counselor about her mood, or to skip shelter appointments.

In the field

Monitoring without explanation can feel like control. For someone who has been controlled, it can make them stop asking for help. Examples from the field?

Trauma-informed ethics require predictability, transparency, and choice.

Conrad (2026); Harris & Fallot (2001); Yu & Rose (2026).

OBJECTIVE 1 · BENEFITS, AND WHERE BIAS ENTERS

Real benefits, and a warning about data

What AI can offer

- Efficiency and more consistent responses
- Earlier awareness of unmet needs, such as housing, income support, childcare, and therapeutic services
- Support for survivor autonomy when aimed at services, not surveillance

Where bias enters

- Administrative data record who gets reported and investigated
- Algorithms learn that history and reproduce it
- High overall accuracy can still hide disparate impact on Black, Indigenous, and low-income families

Chouldechova et al. (2018); Conrad (2026); Dettlaff et al. (2020); Keddell (2019).

APPLY · OBJECTIVE 1 · MAYA'S SCORE

Which harm is most present in the Intake Safety Score?

Restated harms:

Carceral logic: a label authorizes monitoring without support.
Structural bias: system contact and poverty become risk signals.
Epistemic injustice: the score outweighs Maya's account.
Psychological harm: being watched suppresses help-seeking.

Poll

A · Carceral logic
B · Structural bias
C · Epistemic injustice
D · Psychological harm
E · More than one

Then: name the input that drives the most harm for Maya, and why.

LEARNING OBJECTIVE 2

Describe practical strategies for preserving agency, consent, and voice

Survivor agency means voice, transparency, and contestability

OBJECTIVE 2 · AGENCY

Agency is not unlimited autonomy

Definition

The degree to which individuals have meaningful voice, transparency, and contestability within the systems that assess risk and authorize intervention (Lipsky, 1980; Tyler, 2006).

Limits, named honestly

Safety concerns, legal mandates, and cognitive or developmental constraints shape decisions (Herman, 1992; Stark, 2007).

For Maya

She may not control shelter safety decisions. She can still know how the score works, add her account to the record, and contest the result.

Ask of every AI decision: where does the person's voice go?

Conrad (2026), Survivor-Centered AI Framework.

OBJECTIVE 2 · THREE PRINCIPLES

Principles for trauma-informed, antioppressive AI

Principle	Core ethical commitment	What it means in AI design
1 · Agency over automation	Survivor self-determination and consent	AI augments professional judgment and survivor voice; it does not replace them
2 · Bias audits as ethical imperatives	Accountability for structural inequity	Bias is treated as a moral and institutional issue, not only a technical flaw
3 · Redistributive AI	Justice through material support	Prediction is redirected from identifying "risky" people to identifying unmet needs

Conrad (2026), Table 1: Principles for Trauma-Informed, Antioppressive AI in Violence Prevention.

PRINCIPLE 1 OF 3 · AGENCY OVER AUTOMATION

AI must not trigger action on its own

THRESHOLD

AI must not independently trigger investigation, removal, or surveillance without human deliberation and contextual justification.

In Maya's case

Her score of 62 cannot send a report on its own. A worker must weigh it against her account and context.

In practice

Human-in-the-loop review before action. Opt-out mechanisms. Survivor advisory boards reviewing model use (Table 1).

Three conditions follow on the next three slides: consent beyond disclosure, human decisions, and survivor authority.

Conrad (2026), Principle 1; Felzmann et al. (2019); Costanza-Chock (2020).

CONDITION 1 OF 3 · UNDER PRINCIPLE 1

Intelligible consent, beyond disclosure

THRESHOLD

Consent includes an intelligible explanation of how data are used, how outputs influence decisions, and how a person may contest or decline without punitive consequences.

In Maya's case

A form that says "data may be shared with partner agencies" is not enough. Maya needs to know her diagnosis could reach her psychiatrist's chart.

Practice

Give a plain-language summary. Ask Maya to explain it back. Record that she was told she can decline.

Conrad (2026), Principle 1; Felzmann et al. (2019); Selbst et al. (2019).

MAYA ·

What if the score goes into her psychiatrist's record?

RESTATED: CONDITION 1, INTELLIGIBLE CONSENT

Consent goes beyond disclosure. Maya must understand where the score goes, and be able to decline without penalty.

The change

The counseling center shares the intake score with the clinic in the same building. Maya's diagnosis and a "moderate" band now sit in the same chart. She agreed to counseling. She did not agree to share anything with the clinic.

Our answer shifts from "use with care" to "do not share without specific, revocable consent." The harm is now stigma and loss of trust in her treatment.

Question for the room: Does your answer change if the score goes only to the counselor?

CONDITION 2 OF 3 · UNDER PRINCIPLE 1

Human decisions, with the score as one input

THRESHOLD

Final decisions remain with human actors. Algorithmic recommendations inform the institutional response; they do not determine it.

In Maya's case

The worker reviews the score with Maya. The worker records where their own judgment agrees with the score and where it differs, and why.

In practice

The article suggests supervisors facilitate a structured discussion of a risk recommendation: its data, assumptions, and potential harms (Ferguson, 2018; Collins-Camargo & Royse, 2010).

Conrad (2026), Principle 1; Munro (2019); Ferguson (2018).

MAYA ·

What if Maya can see the score, hear what drives it, and ask for review?

The change

The worker explains the score in plain language and shows which inputs contributed most. Maya can ask a supervisor to review the score. She can also decline to have the score used in her plan.

RESTATED: CONDITIONS 1 AND 2

Consent requires an intelligible explanation and the right to contest. The final decision stays with a person, and the score informs it.

About the tool adoption, our answer shifts from "no" to "not yet," provided the review is real and Maya is not required to accept the score. Contestability is what makes agency real.

Question for the room: Is a supervisor review enough, or does Maya need the right to refuse the score entirely?

CONDITION 3 OF 3 · UNDER PRINCIPLE 1

Survivors with real authority over the tool

THRESHOLD

Advisory or cogovernance structures must have authority over how tools are deployed, evaluated, or discontinued. Comment alone is not enough.

In Maya's case

Survivors sit on a board that reviews the Intake Safety Score each quarter, with the authority to pause it.

Practice

Give the board a budget, a formal veto, and access to outcome data. A board that only comments does not meet this threshold.

Conrad (2026), Principle 1 and Public-agency recommendations; Costanza-Chock (2020).

APPLY · OBJECTIVE 2 · MAYA'S INTAKE

Which condition would you fix first?

Restated conditions (Principle 1):

1 · Consent beyond disclosure, with the right to decline.

2 · Human decisions, with the score as one input.

3 · Survivor authority to pause or change the tool.

Poll

A · Consent

B · Human decisions

C · Survivor authority

QUESTIONS ABOUT MAYA'S CASE

What should happen before the score is used?

THRESHOLD, RESTATED (PRINCIPLE 1: AGENCY OVER AUTOMATION)

AI may inform the decision. It may not decide it alone. Maya must keep a meaningful voice, be able to see how the system works, and be able to contest the result.

1. Who should see Maya's score, and who can see it today?

2. Which inputs reflect her partner's abuse, and which reflect poverty, her childhood contact with the system, or her diagnosis?

3. What must Maya know, and be free to say no to, before the worker uses the score?

PRINCIPLE 2 OF 3 · BIAS AUDITS AS ETHICAL IMPERATIVES

A bias audit asks who is labeled, not only how often the model is right

1 · Procedural transparency

Data sources, variables, and performance by demographic group. Model cards and datasheets (Gebru et al., 2021).

2 · Participatory oversight

Practitioners and affected communities help define what counts as harm or fairness (Metcalf et al., 2021).

3 · Remedial accountability

Modify, suspend, or reject tools that produce unjustified disparate impact (West et al., 2022).

In Maya's case

No audit is published. Under this principle, that is a stop sign, not a footnote.

In the field

Obermeyer et al. (2019) shows how an unaudited health algorithm can carry racial disparities into practice.

Conrad (2026), Principle 2; Gebru et al. (2021); Metcalf et al. (2021); West et al. (2022).

MAYA ·

What if the vendor published an independent bias audit?

The change

An independent audit reports accuracy by rurality, by psychiatric diagnosis, and by prior child welfare involvement. It finds the score is least accurate for rural women with mental health diagnoses. The vendor has not changed the model.

RESTATED: PRINCIPLE 2, BIAS AUDITS

Audits are ethical imperatives. They must examine how data are produced and how outputs are read, not only overall accuracy.

Our answer shifts from "no" to "not yet" only if the audit leads to a fix and a review date. An audit that finds a gap and no fix stays at "no." Accurate on average does not mean fair.

Question for the room: Would an audit that reports only overall accuracy change your answer?

PRINCIPLE 3 OF 3 · REDISTRIBUTIVE AI

Point prediction at resources, not at people to watch

THE QUESTION

What would it mean for predictive capacity to be oriented toward redistribution rather than surveillance?

What it asks for

Forecast housing instability, childcare needs, or service deserts. Build detection that triggers care rather than punishment.

In Maya's case

The tool would flag her need for childcare, housing, and coordinated psychiatric care, instead of a risk band that brings scrutiny.

If the agency cannot provide what a prediction points to, deploying the prediction may be ethically problematic. Refusal may be the most consistent response.

Conrad (2026), Principle 3; Eubanks (2018); Mullaly (2010).

MAYA ·

What if a "high" band triggers a child welfare report, not support?

The change

In this county, a high band sends a report to child welfare automatically. Shelter advocacy, housing help, and childcare are offered only if Maya asks for them. Maya was in the child welfare system herself as a child.

RESTATED: PRINCIPLE 3, REDISTRIBUTIVE AI

Ask what the tool connects people to. A prediction should point to resources, not only to monitoring.

Our answer shifts from "use with support" to "do not use this design." A report without added help makes the tool monitoring, and it brings scrutiny of Maya as a parent.

Question for the room: Would you use the tool if a high band added advocacy, housing, and childcare, but no report?

LEARNING OBJECTIVE 3

Apply an evaluation framework before adopting a tool

For tools used with vulnerable and trauma-exposed people

OBJECTIVE 3 · BEFORE ADOPTION

The threshold test: three conditions must hold

(a) It does not intensify surveillance without corresponding support.

(b) It does not substitute automation for relational labor.

(c) It does not displace professional judgment or survivor voice.

If these cannot be met, delay, constraint, or refusal may be the most trauma-informed response.

OBJECTIVE 3 · EVALUATION FRAMEWORK

Check before, during, and after

Before: impact assessment

Document purpose, data sources, decision stakes, affected groups, and potential harms. Ask whether the tool is appropriate given capacity, consent, and power.

During: bias audit

Require documentation and performance by group. Define harm with practitioners and affected people. Fix or suspend tools with unjustified disparities.

After: debrief and govern

Review performance, unintended effects, and survivor feedback routinely. Keep the authority to pause or reject the tool.

Success is measured by reduced inequity, better access to care, and survivor-defined safety, not accuracy alone.

Conrad (2026), Implications for Practice, Policy, and Research; Reisman et al. (2018); Gebru et al. (2021); Metcalf et al. (2021).

BEFORE ADOPTION · IMPACT ASSESSMENT

Document what the tool does before anyone uses it

Document

Purpose. Data sources. Decision stakes. Affected populations. Potential harms. Institutional capacity to respond. Consent conditions. Power asymmetries.

Maya's tool, first pass

Stakes: shelter priority and reporting. **Data:** child welfare history, benefits, housing, diagnosis. **Gap:** the vendor cannot say how a diagnosis is weighted.

Restated threshold: an unexplained weighting of a diagnosis is a reason to pause, not to proceed.

Reisman et al. (2018); Conrad (2026), Social Work Policy: Regulating AI Through Ethical Thresholds.

DURING USE · BIAS AUDIT

Check who is labeled, group by group

Documentation

A model card or datasheet. Data sources and limits (Gebru et al., 2021).

Performance by group

Rural residents, people with psychiatric diagnoses, people with child welfare history, and race and income where data allow.

Maya's check

Ask for error rates for rural women with diagnoses, and for people with prior child welfare involvement. Ask what changed after each finding.

Restated threshold: a disparity that is unjustified and unfixed means the tool is suspended or rejected.

Conrad (2026), Principle 2; Gebru et al. (2021); Metcalf et al. (2021); West et al. (2022); Obermeyer et al. (2019).

AFTER DEPLOYMENT · DEBRIEF AND GOVERN

Keep the power to pause, and track what happens next

Routine debriefs

Frontline workers, supervisors, technologists, and advocates review performance and unintended effects, with survivor feedback.

Track outcomes

For every high band: did it lead to a report, to services, or to nothing? Track who received which response.

Maya's six-month review

How many high bands led to advocacy and housing help, and how many led to reports? Is any group getting more reports than services?

Authority to pause, modify, or reject must sit with a body that includes survivors.

Conrad (2026), Social Work Practice and Social Work Policy sections; D'Ignazio & Klein (2020); Knight (2020).

APPLY · OBJECTIVE 3 · MAYA'S TOOL

Does Maya's tool pass the threshold test?

Restated threshold test:

(a) Does not intensify surveillance without corresponding support.
(b) Does not substitute automation for relational labor.

(c) Does not displace professional judgment or survivor voice.

Maya's facts: a score of 62 sets shelter priority and can trigger reporting. The vendor has not published subgroup results or a bias audit.

Poll

Yes, it passes

Not yet: it passes only with changes

No, it fails

Then: which one condition would you test first, and what evidence would change your answer?

WHAT YOU CAN DO NEXT WEEK

Practical steps for your setting

- Ask any vendor for documentation: purpose, data sources, and performance by group.
- Build algorithmic literacy into supervision: examine a score's data and assumptions before accepting it.
- Propose an ethics review step before adoption, similar to an institutional review board.
- Schedule routine debriefs on tool performance, unintended effects, and survivor feedback.

Conrad (2026), Social Work Practice and Social Work Policy sections.

TAKEAWAYS

Three things to carry forward

1 · Agency is voice, transparency, and contestability. Not unlimited autonomy, but a real ability to understand and challenge decisions.

2 · Bias enters through data and structure. Auditing who is labeled, and how well, is ethical work, not only technical work.

3 · Evaluate before you adopt, and be ready to refuse. If the threshold conditions cannot be met, delay or refusal may be the trauma-informed choice.

CHAT · 4 MINUTES

Which information mattered most to you?

What is one thing you learned today that you will take with you?

Read-out · 2 minutes

We will read three answers aloud. Responses are anonymous unless you choose to be named.

QUESTIONS AND DISCUSSION

Q&A

Type your question in chat. Thank you for your time.

Aislinn Conrad, PhD · University of Iowa School of Social Work



SunCloud Health

Supporting Your Practice

To collaborate and learn more about our programs:

Autumn Aumann
Director of Outreach
SunCloud Health
aaumann@suncloudhealth.com
773.391.0407