RESEARCH AND BUILD PROPOSAL

Support, Not a Label

A Transparent, Open-Source Early Warning Model for Virginia Students

Introducing Compass, a student early warning model built to be explainable, auditable, and used only to offer support

MANGO (Modern Advocates for New Global Opinions)
Neel Shah, Anish Ranga, Jeev Dalawari, Ishant Mekala
Richmond, Virginia
October 2026

I. Executive Summary

Many American schools use early warning systems to predict which students are at risk of not graduating. The idea is sound: find students who are struggling early, while there is still time to help. The execution has not always matched the idea.

Wisconsin offers the clearest warning. For about a decade the state ran a statewide algorithm that scored middle schoolers using dozens of data points, including race and family income. A 2023 investigation by The Markup found that when the system predicted a student would not graduate on time, it was wrong nearly three quarters of the time, and it was wrong far more often for Black and Hispanic students than for white students. A large academic study of the same system found that its labels did not improve graduation rates for the students it flagged.

MANGO believes the answer is not to abandon early warning but to build it better, in the open, where anyone can check the work. This paper introduces **Compass**, an open-source early warning model that MANGO students are building. Compass is designed around six commitments:

1. **No demographic inputs.** Compass never scores students using race, ethnicity, sex, income, disability, English learner status, or home address.
2. **Every flag explains itself.** Each result shows, in plain language, which attendance, behavior, or course indicators triggered it.
3. **Fairness is measured, not assumed.** Compass reports error rates for every student group and publishes the results.
4. **Support, not labels.** Compass suggests check-ins. It never labels a student "high risk" and is never used for discipline or placement.
5. **Student data stays in the school.** Compass runs inside a school division. MANGO never receives student records.
6. **Open source.** The code, documentation, and evaluation results are public.

We are honest about what a prototype can and cannot show. Compass is being developed and tested on synthetic data and a public research dataset. We will not claim it works for Virginia students until it has been validated with real Virginia data through a formal agreement with a school division. This paper also recommends guardrails that any Virginia division using a predictive early warning system should adopt, whether or not it ever uses Compass.

II. Why Early Warning Systems Exist

Students rarely drop out without warning. Decades of research show that the clearest signals appear in three areas, often called the ABC indicators: **A**ttendance, **B**ehavior, and **C**ourse performance. A student who misses a large share of school, receives repeated suspensions, or fails a core course in ninth grade is much less likely to graduate on time.

Early warning systems turn those signals into a list that counselors and teachers can act on. By 2015, more than half of U.S. public schools had implemented some version of one, according to a U.S. Department of Education survey cited by Perdomo et al.

Virginia has its own tool. The Virginia Department of Education developed the Virginia Early Warning System (VEWS) with four school divisions and national partners. It uses data already held by schools to

identify ninth graders who are off track to earn a diploma and to help divisions direct resources before students drop out. VEWS is built on simple, readable thresholds. For example, missing 10 percent or more of instructional time is treated as a red flag, which matches the state's definition of chronic absenteeism.

Simple indicator systems like VEWS have a real strength: everyone can see why a student was flagged. The newer trend is toward machine learning systems that combine many more variables. Those systems promise more accuracy but are harder to inspect, and that is where the problems in Wisconsin began.

III. What Went Wrong in Wisconsin

From 2012 until 2023, the Wisconsin Department of Public Instruction ran the Dropout Early Warning System (DEWS). Twice a year, schools received a list of students in grades 6 through 9 with a color-coded label: green for low risk, yellow for moderate, red for high. According to The Markup, DEWS was an ensemble of machine learning models trained on historical data that included test scores, discipline records, free or reduced-price lunch status, and race.

A. The false alarms fell unevenly

An internal 2021 equity analysis by the state, obtained by The Markup through a public records request, found that the system's false alarm rate was 42 percentage points higher for Black students than for white students and 18 percentage points higher for Hispanic students. A false alarm here means a student the system predicted would not graduate on time who in fact did. The slide summarizing that analysis was titled "Is DEWS Fair?" and its notes concluded that it was not.

Part of this was by design. A 2015 paper describing DEWS stated that the agency was willing to accept 25 false alarms for every dropout it correctly identified. A University of Pennsylvania researcher who studies these systems told reporters that using demographic factors as predictors tends to overemphasize those variables and produce exactly this kind of error.

B. The labels did not improve outcomes

Researchers from UC Berkeley and other institutions studied nearly a decade of DEWS data covering more than 200,000 students. Their findings, published at the 2025 ACM Conference on Fairness, Accountability, and Transparency, are more surprising than a simple story of a broken algorithm. The system did accurately sort students by risk. But being labeled high risk did not lead to higher graduation rates.

The researchers also showed that information about a student's school and district, without any individual risk score, could target support about as efficiently as the full individual prediction system. Their conclusion is that dropout is largely driven by structural factors, so accurate individual predictions add less value than people assume.

C. Students paid a hidden cost

Students interviewed by The Markup were surprised to learn the system existed and said the high-risk labels felt stigmatizing and discouraging. Two Wisconsin students later wrote that their school system should focus on helping students who need real support instead of predicting failure.

IV. The Lesson: Prediction Is Not the Hard Part

Wisconsin teaches three lessons that shape everything about Compass.

1. **Accuracy is not fairness.** A model can rank students well overall and still make far more mistakes for some groups than others. Fairness has to be measured group by group.
2. **A label is not an intervention.** A risk score only helps if it leads to a specific, effective support. Without that, it is a label that follows a student around.
3. **Context matters as much as the individual.** If needs are concentrated in certain schools, the first question should be whether those schools have enough counselors, tutors, and resources, not which individual child to flag.

These lessons point to a different design goal. Instead of asking "how do we predict dropout as accurately as possible," Compass asks "how do we help schools notice students who need a check-in, explain why, and prove the process treats every group fairly?"

V. What We Are Building: Compass

Compass is an open-source software package that a school division can run on its own systems. The table below summarizes how its design responds to the problems documented in Wisconsin.

Issue	Wisconsin DEWS	Compass
Inputs	Over 40 features, including race and free or reduced-price lunch status	Attendance, behavior, and course performance only. No demographic inputs.
Output	Low, moderate, or high risk label	"Check-in suggested" with the specific reasons listed
Explainability	Ensemble model, hard to inspect	Interpretable model. Every factor and its weight is published.
Fairness checks	Internal analysis, made public through a records request	Built-in audit of error rates by group, published with every release
Data location	State system	Runs inside the division. MANGO never sees student records.
Code	Not public	Open source

A. No demographic inputs

Compass scores students only on things a school can act on: attendance, behavior incidents, course grades, failures, and credits earned. It never uses race, ethnicity, sex, family income, disability status, English learner status, or address. Removing these variables does not by itself guarantee fairness, because other data can act as a stand-in for them. That is why the next commitment exists.

B. Fairness is measured, not assumed

Demographic data is used in exactly one place: the audit. After the model makes its predictions, Compass compares error rates across student groups, including false alarm rates, miss rates, and calibration. Those audit results are published with every version. If a version shows a meaningful gap in false alarms between groups, it does not ship until the gap is reduced or clearly explained.

We also acknowledge a known limit. Research by Kleinberg, Mullainathan, and Raghavan and by Chouldechova shows that when groups have different underlying rates of an outcome, no risk score can

satisfy every definition of fairness at once. Compass therefore makes its priority explicit: **because a false alarm can stigmatize a student, Compass prioritizes equal false alarm rates across groups**, and it reports the tradeoffs this creates.

C. Every flag explains itself

Compass uses an interpretable model, such as a points-based scorecard or a constrained logistic regression, rather than a black box. When a student is flagged, the counselor sees the reasons in plain language, for example: "Missed 14 percent of school days this quarter" and "Failed Algebra I first semester." Because every indicator is something the school can respond to, every flag points toward a specific kind of support.

D. Support, not labels

Compass never produces a "high risk" label. It produces a short list of students for whom a check-in is suggested, along with the reasons. Its documentation and license terms state that results may not be used for discipline, course placement, program eligibility, or any record that follows a student. Flags expire at the end of each grading period and are recalculated from current data.

E. School-level view first

Following the Wisconsin research, Compass includes a school-level dashboard that shows where needs are concentrated across a division. This helps leaders ask whether certain schools need more counselors or resources before focusing on individual students.

F. Privacy by design

Compass runs entirely on division systems. MANGO does not collect, receive, or store any student records. The public demonstration on MANGO's website uses only synthetic students who do not exist.

VI. How Compass Works

The table below lists the inputs Compass uses. Each is data Virginia schools already collect.

Indicator	Examples of what is measured	Why it matters
Attendance	Share of days missed this grading period and year to date	Absenteeism is a well-documented predictor of dropout. Virginia defines chronic absence as missing 10 percent or more of the year.
Behavior	Count of suspensions and serious incidents	Repeated removal from class means lost instruction time.
Course performance	Core course failures, GPA, credits earned toward on-time promotion	Ninth grade course failure is one of the strongest early signals.
Trend	Change in the above since the previous grading period	A sudden drop can signal a new problem worth a conversation.

The model produces a score between 0 and 1. A division chooses how many check-ins its staff can realistically handle, and Compass suggests that many students, ranked by need, with reasons attached. Compass is evaluated on several measures, because no single number tells the whole story:

- **Ranking quality**, measured with standard metrics such as the area under the ROC curve.

- **Precision at capacity**, meaning how many suggested check-ins are students who truly needed support, at the number of students staff can actually see.
- **Calibration**, meaning whether a score of 0.3 really corresponds to about a 30 percent chance, for every group.
- **Group error rates**, meaning false alarm and miss rates for each student group.

To make the effect of design choices visible, Compass is always compared against a baseline model that does use demographic inputs, similar in spirit to DEWS. Publishing that comparison shows exactly what is gained and what, if anything, is lost by leaving demographics out.

VII. Data and Validation Plan

MANGO is a student organization and does not have, and should not have, access to real student records. Compass is therefore being developed in three phases.

Phase 1: Synthetic data

We generate a population of synthetic students whose attendance, behavior, and grades follow patterns described in published research. We deliberately build in the kind of structural inequality seen in real data, such as some groups attending under-resourced schools, so that we can test whether Compass avoids the proxy problems that hurt DEWS. All results from this phase are labeled as synthetic.

Phase 2: Public research data

We test Compass on a real but public dataset: the UCI Student Performance dataset, which covers 649 secondary students at two schools in Portugal and is licensed for reuse under CC BY 4.0. It is small and not from Virginia, so it serves as a check on real-world behavior, not proof that Compass works here.

Phase 3: Validation with a Virginia division

The final step is a partnership with a Virginia school division. Under a formal written data agreement consistent with FERPA, such as the exception for studies conducted for a school (34 C.F.R. § 99.31(a)(6)), the division would run Compass on its own historical, de-identified records and share only aggregate results. MANGO would publish those results, including any failures. Until this phase is complete, MANGO will describe Compass as a prototype.

VIII. Limits and Risks

We believe it is more credible to name the risks of our own project than to hide them.

- **A model cannot fix under-resourced schools.** If a school has one counselor for hundreds of students, a better list does not create more time. Compass's school-level view is meant to make that problem visible, not to replace investment.
- **Proxies can still cause bias.** Even without demographic inputs, attendance and discipline records can reflect unequal treatment. That is why the audit exists and why results are published.
- **People may trust the score too much.** Staff can over-rely on a number. Compass shows reasons instead of labels and requires a human conversation, not an automatic action.

- **Tools can be misused.** Open-source code can be modified. Compass's documentation and license terms prohibit disciplinary or placement uses, and our policy recommendations ask the state to prohibit them as well.
- **We are students.** We are building this with care, but Compass should be reviewed by professional data scientists, educators, and legal advisors before any division relies on it.

IX. Recommendations for Virginia

Whether or not any division ever uses Compass, MANGO recommends that the Virginia Department of Education issue guidance requiring that any predictive early warning system used by a Virginia school division meet the following standards:

1. **No protected characteristics as inputs.** Race, ethnicity, sex, disability, English learner status, and family income may be used only to audit a model, never to score a student.
2. **Transparency.** Divisions publish which system they use, what data it uses, and how flags are generated.
3. **Annual fairness audit.** Divisions or vendors report false alarm and miss rates by student group each year.
4. **Notice to families.** Students and parents are told that an early warning system is in use and how to ask questions about it.
5. **Human review and support only.** Flags may prompt a conversation or support, never discipline, course placement, or a permanent record.
6. **Measure the help, not just the prediction.** Divisions track whether the supports offered to flagged students actually improve outcomes.

X. Timeline

When	Milestone
Fall 2026	Build the Compass prototype and synthetic data generator. Publish the code and this paper.
Winter 2026 to 2027	Complete Phase 2 testing on public data. Publish the first fairness audit and model documentation.
Spring 2027	Seek a Virginia school division partner for Phase 3 validation. Share recommendations with the Virginia Department of Education.

MANGO began because we wanted student ideas to leave the conference room. Compass is that idea in practice: students identifying a real problem, studying how others got it wrong, and building something better in public, where anyone can check our work.

Sources

- Todd Feathers, "False Alarm: How Wisconsin Uses Race and Income to Label Students 'High Risk'," The Markup, April 27, 2023. themarkup.org
- The Markup, "Takeaways from Our Investigation into Wisconsin's Racially Inequitable Dropout Algorithm," April 27, 2023. themarkup.org
- The Markup, "We're Not Living a 'Predicted' Life: Student Perspectives on Wisconsin's Dropout Algorithm," December 21, 2023. themarkup.org
- Juan Carlos Perdomo, Tolani Britton, Moritz Hardt, and Rediet Abebe, "Difficult Lessons on Social Prediction from Wisconsin Public Schools," Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. doi.org/10.1145/3715275.3732175
- Virginia Department of Education, Virginia Early Warning System (VEWS). VDOE
- Virginia Department of Education, Attendance and School Engagement. doe.virginia.gov
- Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores," 2016. arxiv.org/abs/1609.05807
- Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," Big Data, 2017. arxiv.org/abs/1703.00056
- Paulo Cortez, Student Performance [Dataset], UCI Machine Learning Repository, CC BY 4.0. archive.ics.uci.edu
- Family Educational Rights and Privacy Act regulations, 34 C.F.R. § 99.31(a)(6).