

# Mission-Critical Evals at Scale

Learnings from 100,000 medical decisions

Dr Christopher Lovejoy, MD

Anterior



# Anterior

We're a **New York-based, clinician-led company**, using responsible AI to tackle the burden of healthcare administration

BACKED BY



and FOUNDERS OF



ALUMNI OF





# It's *easy* to go from 0 -> MVP

(and it's getting even easier)



**But what about MVP -> serving  
customers at scale?**



**Q: Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?**





# Q: Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?

 Florence:

The medical record shows a brain MRI from 17/5/2022 that demonstrates 'multiple foci of T2/FLAIR hyperintensity in the infratentorial, juxtacortical and periventricular white matter' which is noted to be 'consistent with multiple sclerosis.' This confirms prior brain MRI findings suspicious for MS.



# Q: Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?

 Florence:

The medical record shows a brain MRI from 17/5/2022 that demonstrates 'multiple foci of T2/FLAIR hyperintensity in the infratentorial, juxtacortical and periventricular white matter' which is noted to be 'consistent with multiple sclerosis.' This confirms prior brain MRI findings suspicious for MS.





So how do you identify and handle  
failure cases?



PART ONE

# Human Reviews of AI Outputs



Mary Seacole

Elephant ▾



Inbox

3

WorkFlos

FloNotes

Policies

Apps

Workspace Data



Cases

Members

Scalpel

Admin

Help Center

Anterior

< Clinical Tasks

## MRI Cervical Spine

95782

Activity

Outcomes

Review



Procedure

MRI Cervical Spine

Outcome Path

Ataxic Gait

▼ ✓ Approval

▼ ✓ Ataxic gait due to neurological issue, as indicated by 1 or more of the following:

✓ Prior brain MRI suspicious for Multiple Sclerosis



Question 1/1

Current Question

Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?

Answer

Yes. The medical record shows a brain MRI from 17/05/2022 that demonstrates 'multiple foci of T2/FLAIR hyperintensity in the infratentorial, juxtacortical and periventricular white matter'. <sup>p. 2</sup> which is noted to be 'consistent with multiple sclerosis.' <sup>p. 3</sup> This confirms prior brain MRI findings suspicious for MS.

If incorrect, please explain why...

Correct

Incorrect

+ TAG FAILURE MODE

Scalpel



Evidence.pdf

Guidelines for CPT Code 95782.pdf

Page 1/10

1.0x



Mary Seacole

Elephant ▾



Inbox

3

WorkFlos

FloNotes

Policies

Apps

Workspace Data



Cases

Members

Scalpel

Admin

Help Center

Anterior

< Clinical Tasks

Scalpel



## MRI Cervical Spine

95782

Activity

Outcomes

Review



Procedure

MRI Cervical Spine

Outcome Path

Ataxic Gait

▼ ✓ Approval

▼ ✓ Ataxic gait due to neurological issue, as indicated by 1 or more of the following:

✓ Prior brain MRI suspicious for Multiple Sclerosis



Question 1/1

Current Question

Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?

Answer

Yes. The medical record shows a brain MRI from 17/05/2022 that demonstrates 'multiple foci of T2/FLAIR hyperintensity in the infratentorial, juxtacortical and periventricular white matter'. 

p. 2

 which is noted to be 'consistent with multiple sclerosis.' 

p. 3

 This confirms prior brain MRI findings suspicious for MS.

Elsewhere in the medical record, it states the patient has confirmed MS - so the MRI is not 'suspicious' for MS - the patient is known to have it

Correct

Incorrect

+ TAG FAILURE MODE

Evidence.pdf

Guidelines for CPT Code 95782.pdf

Page 1/10

1.0x



(

**question** = “Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?”

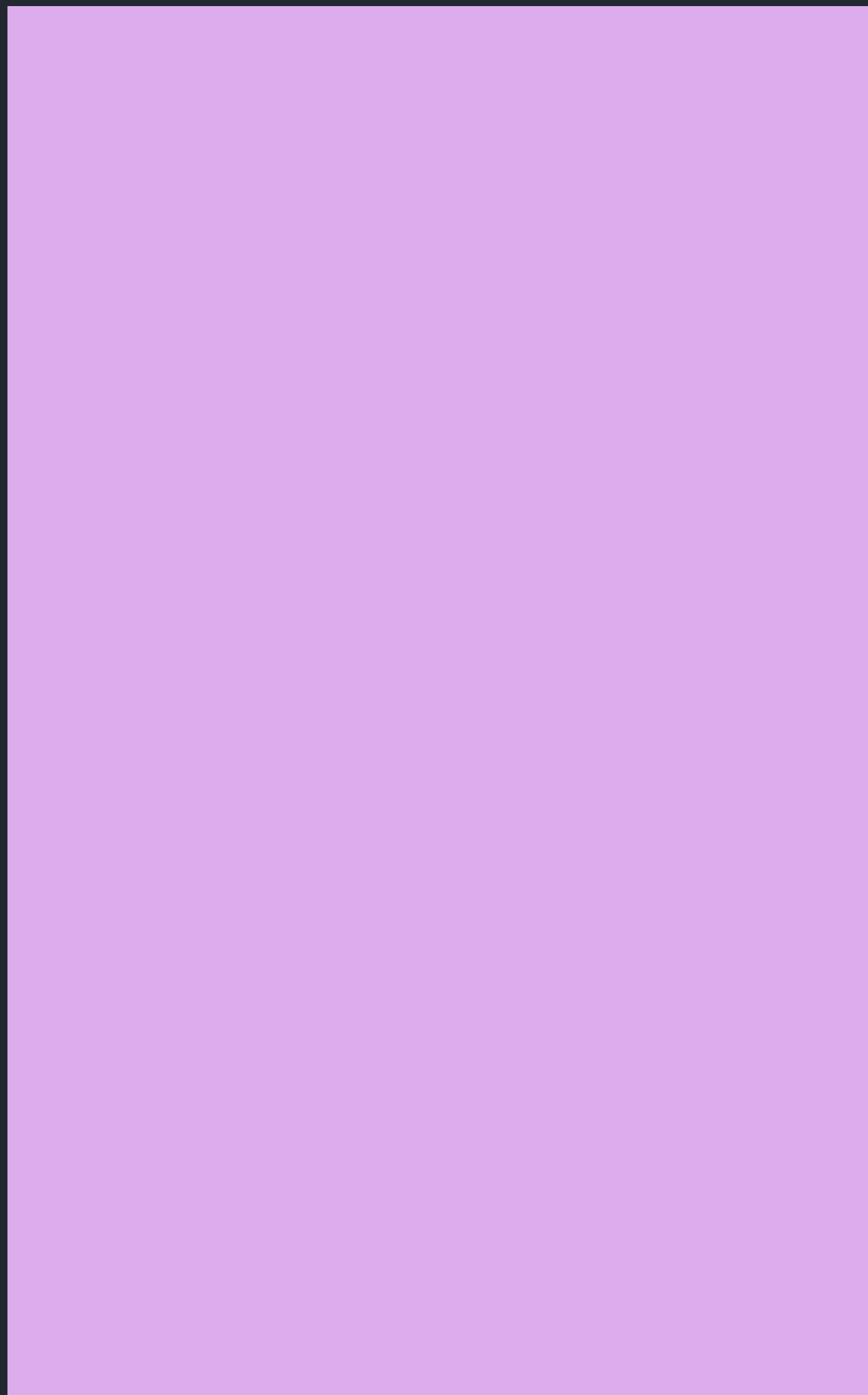
**is\_met** = False

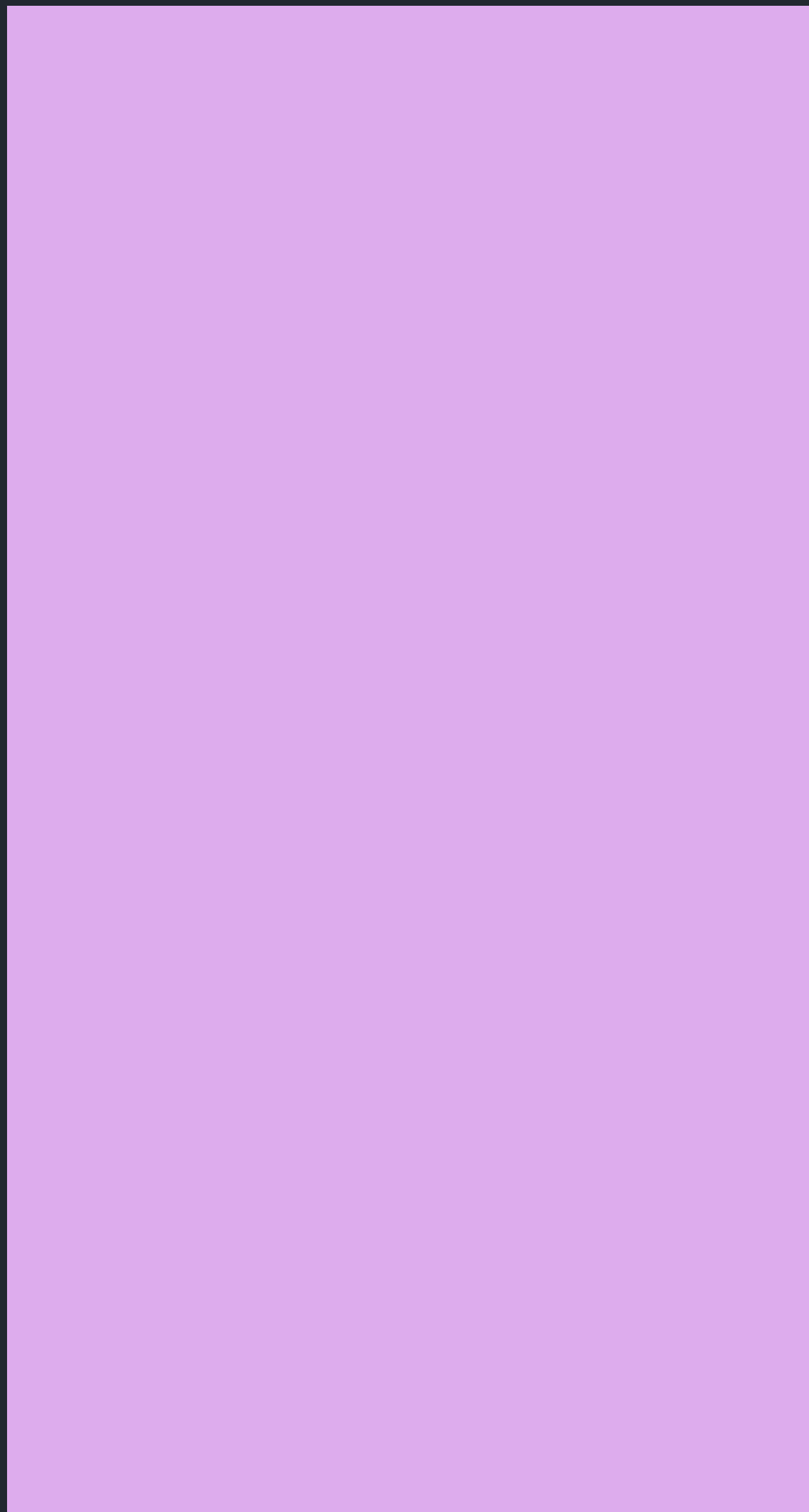
**reasoning** = “While there is a brain MRI from 17/5/2022 described as ‘consistent with multiple sclerosis’, this finding is documented in the context of established Multiple Sclerosis diagnosis. The medical record confirms a definitive MS diagnosis, precluding classification of these imaging findings as ‘suspicious for’ MS. Rather, these MRI findings represent characteristic radiological features consistent with the patient's known MS diagnosis.”

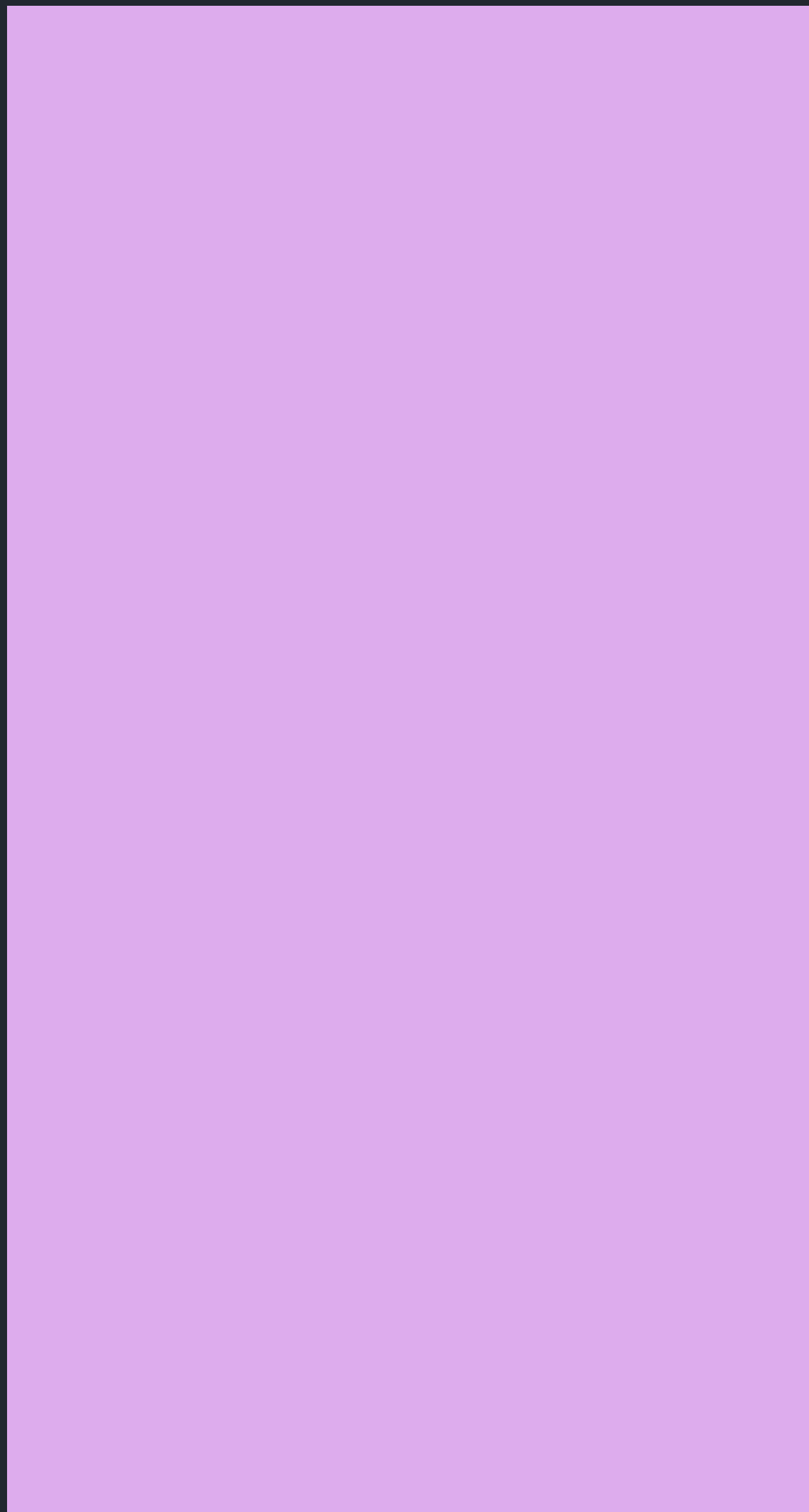
)

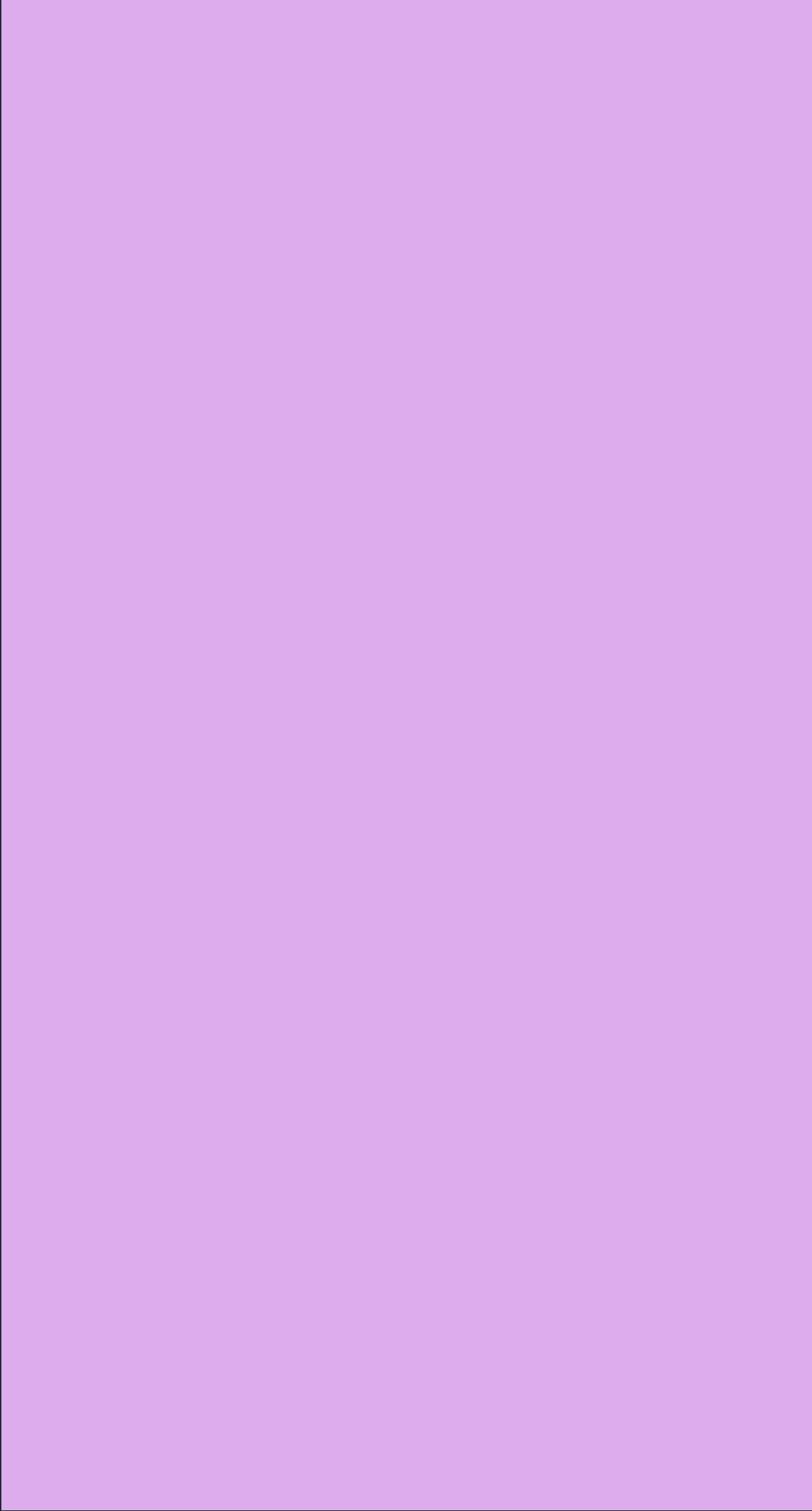
1,000 MEDICAL  
DECISIONS / DAY

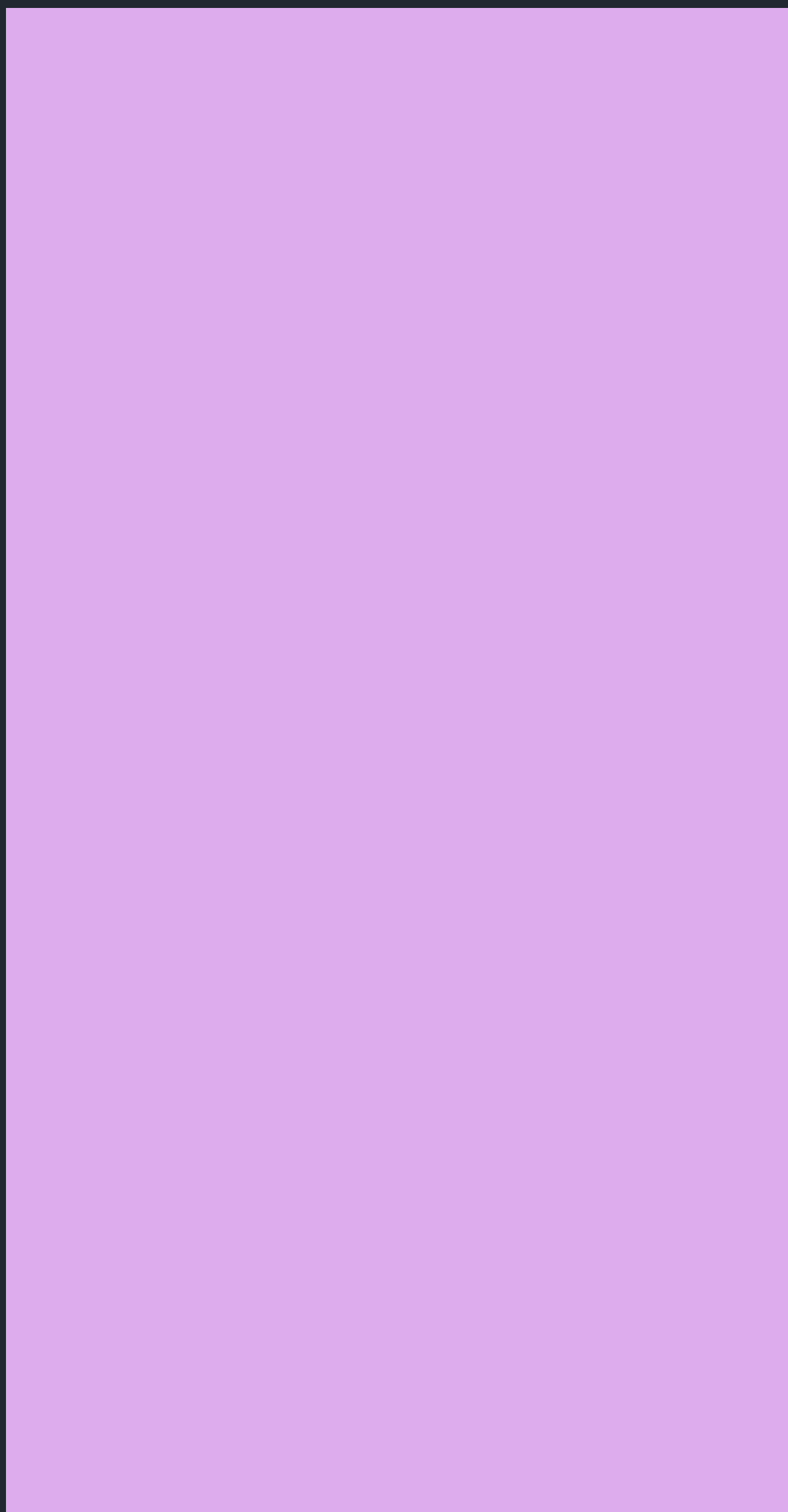
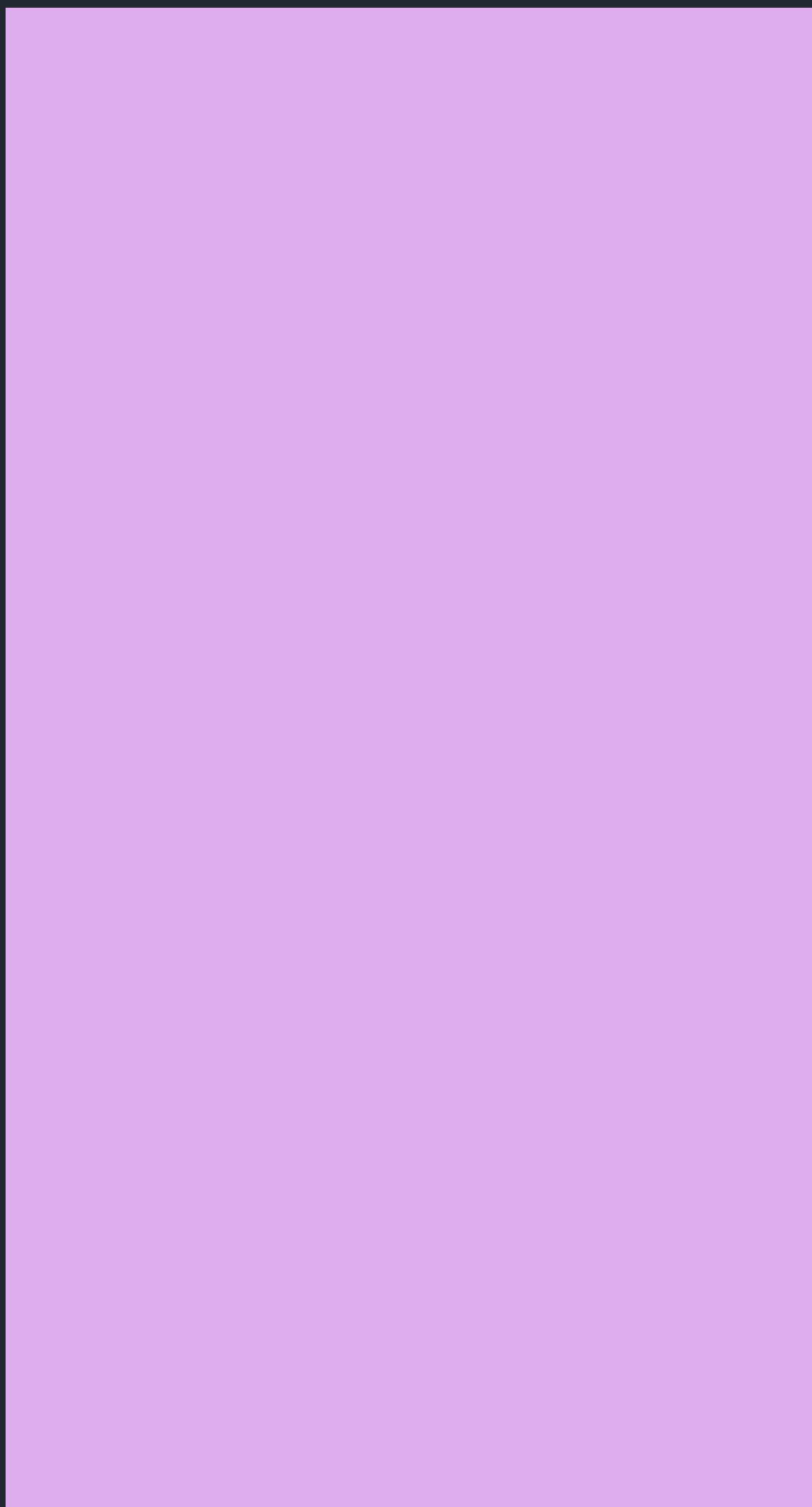


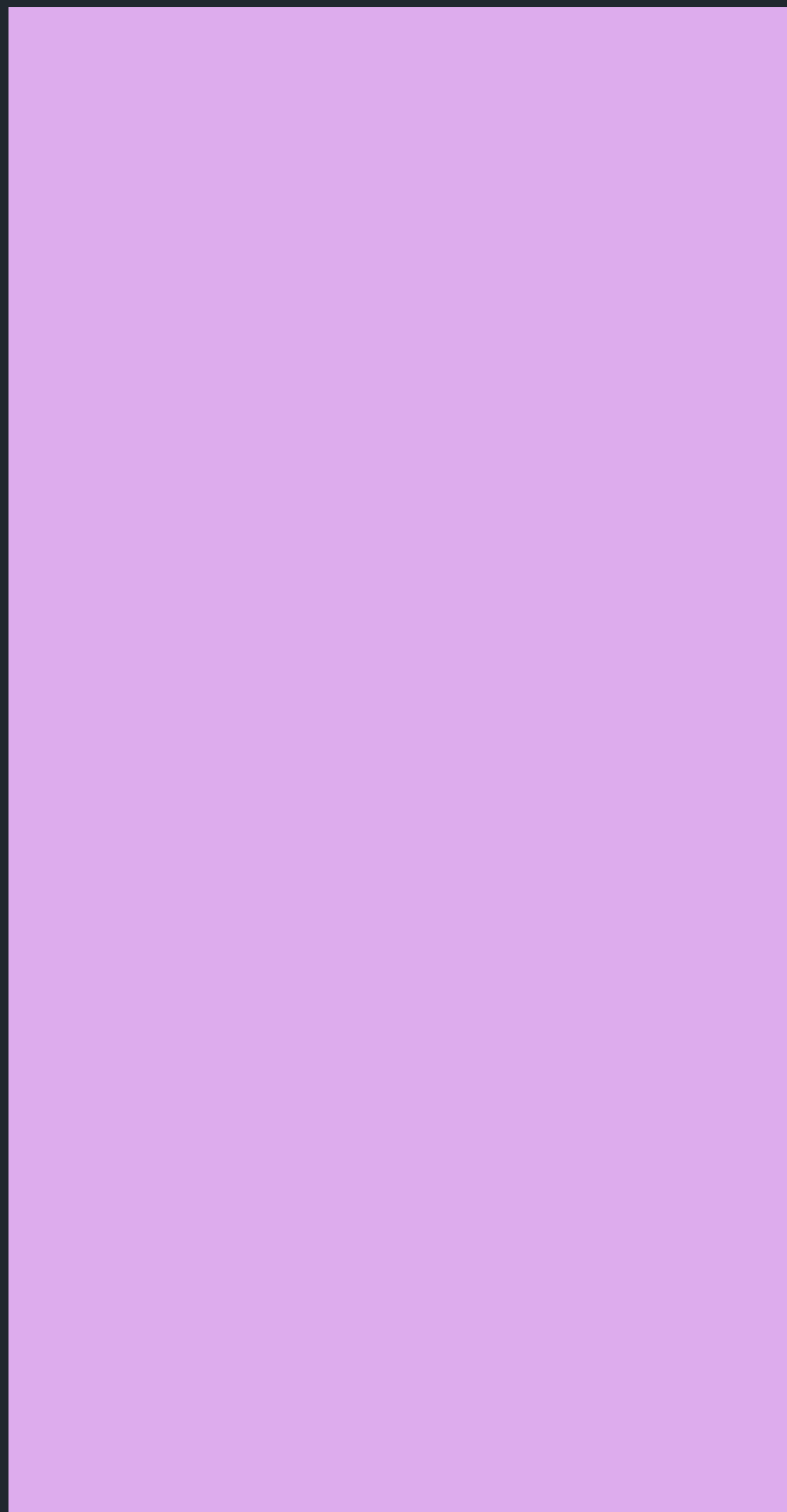
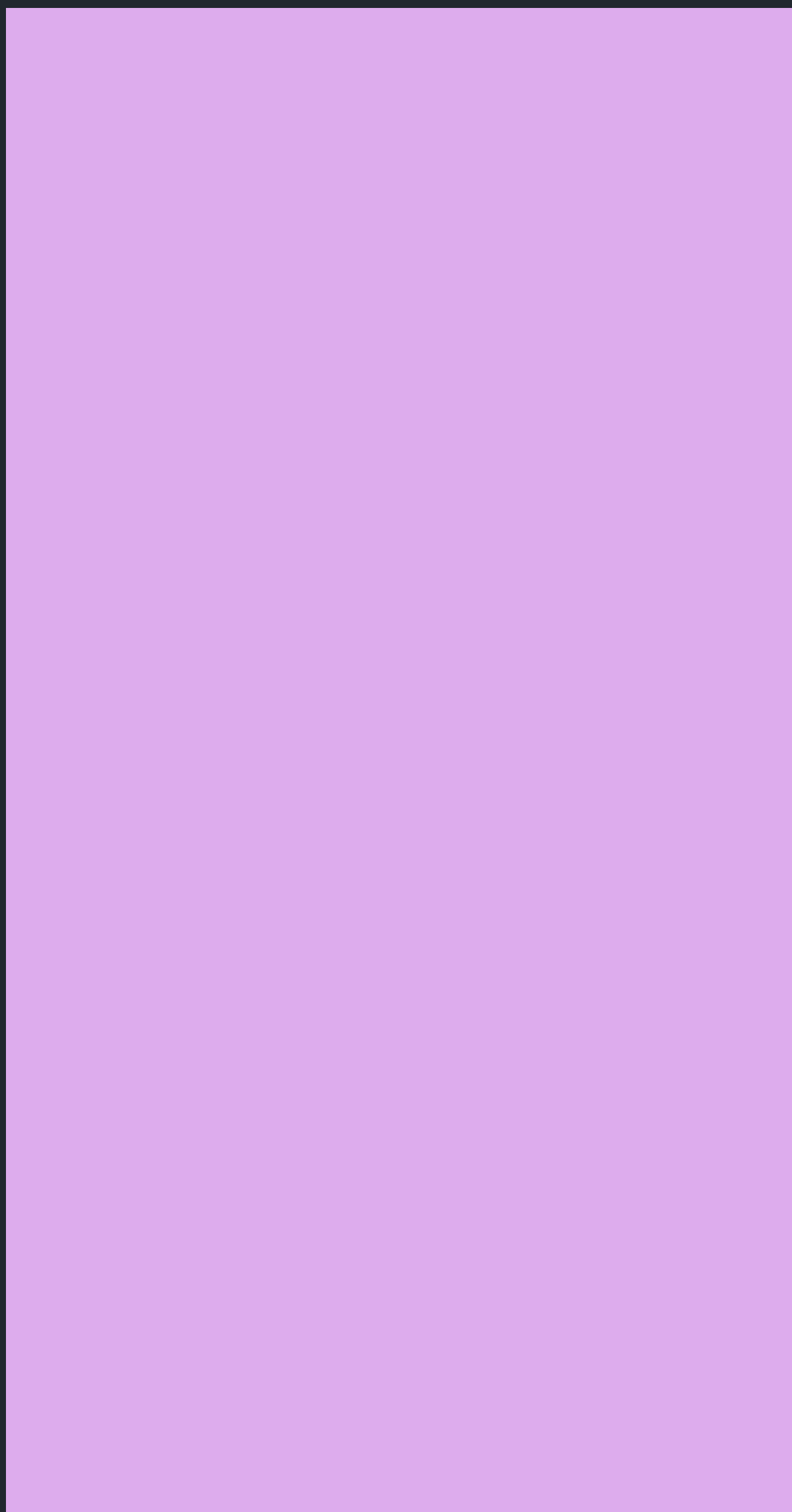
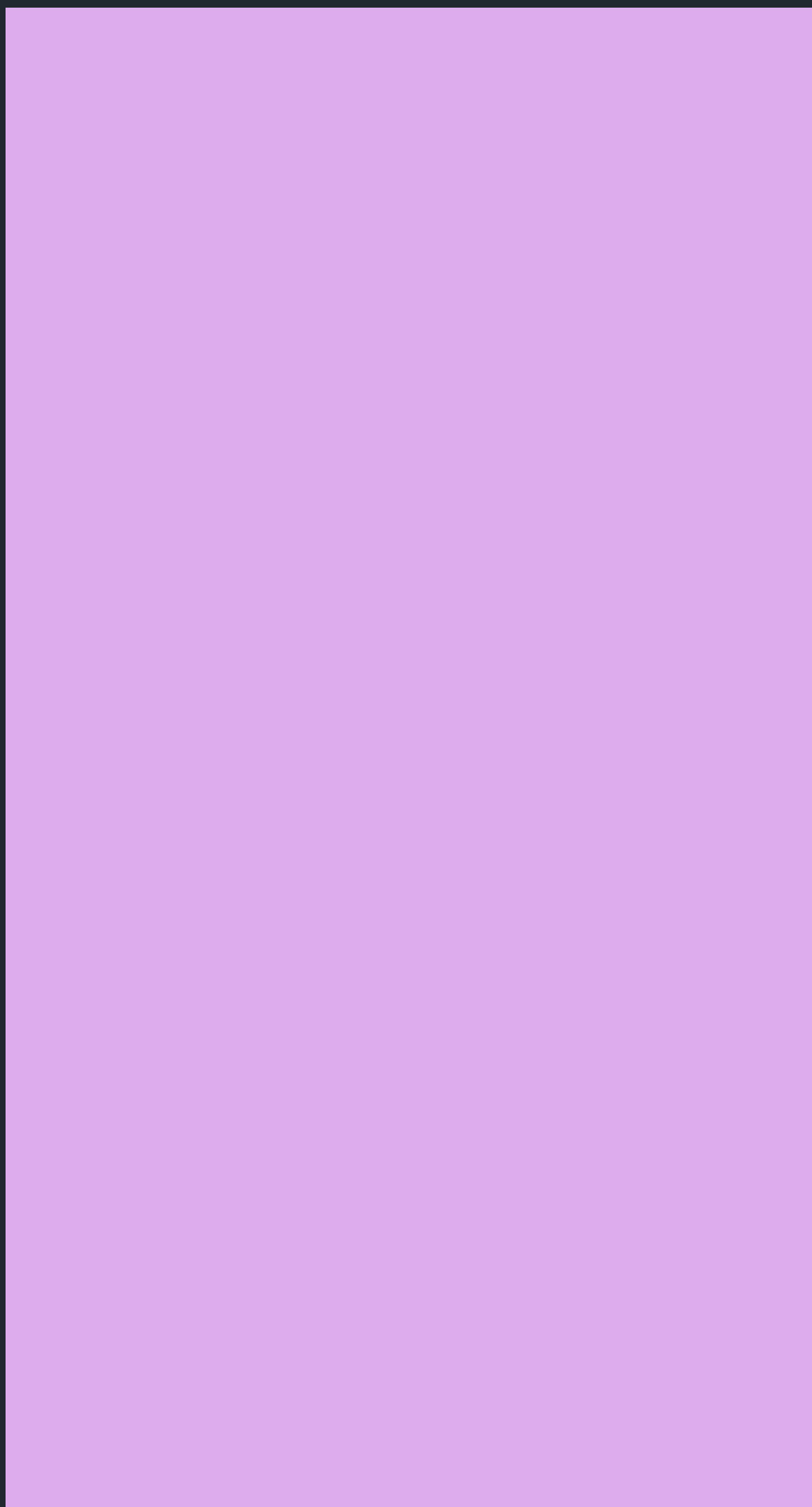
















The problem: this doesn't scale.





PART TWO

# Offline Eval Datasets





Relying only on offline evals is  
playing with fire



THE SOLUTION

**A real-time reference-free  
evaluation system**



# LLM-as-judge is a great starting point





Our reference-free evals generate both  
**confidence score** and **predicted output**



Predicted output gives us estimated  
performance metrics

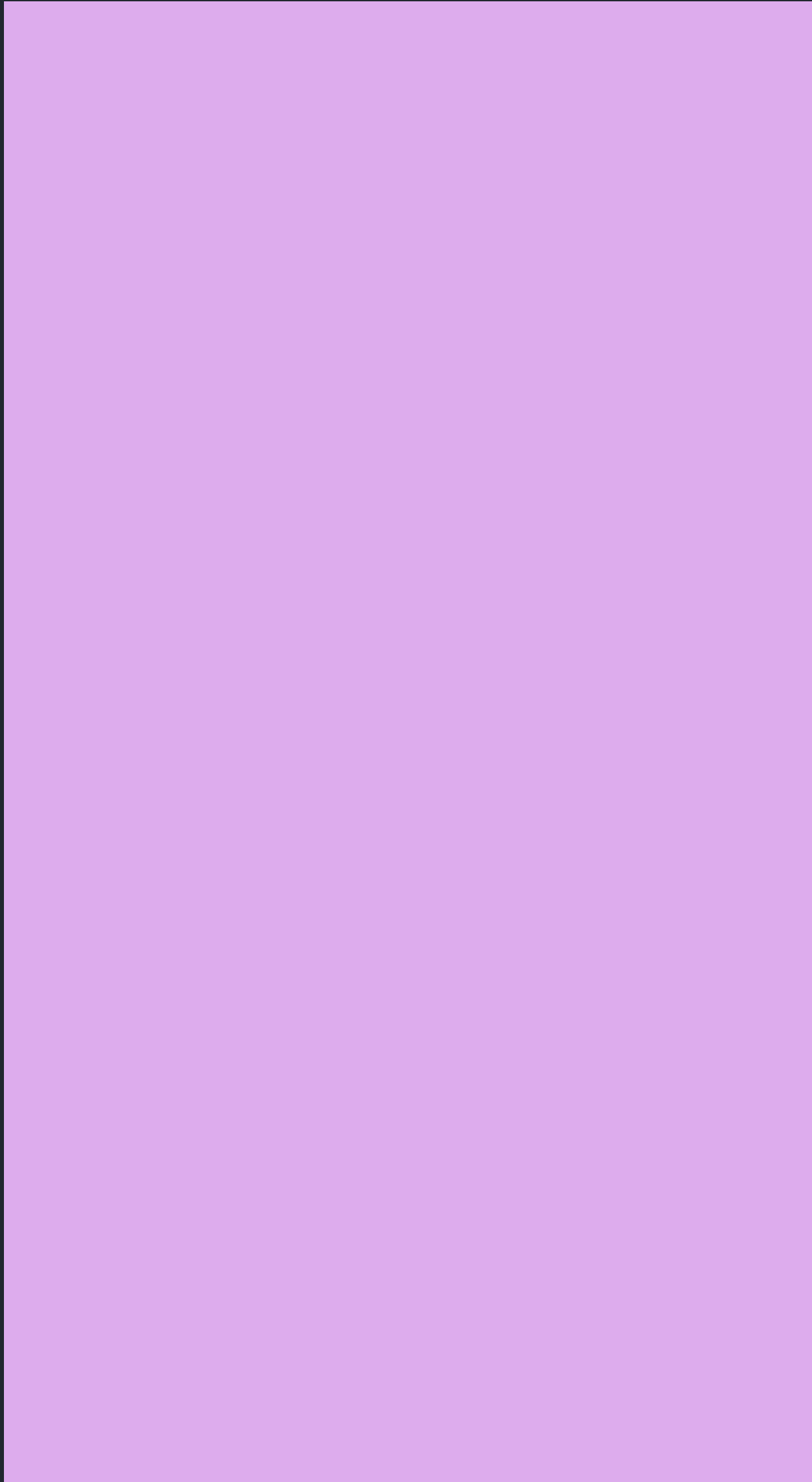




**Predicted output** gives us estimated performance metrics... and alignment scores



Confidence grading can be used to prioritise reviews 



Confidence grading can be used to   
prioritise reviews



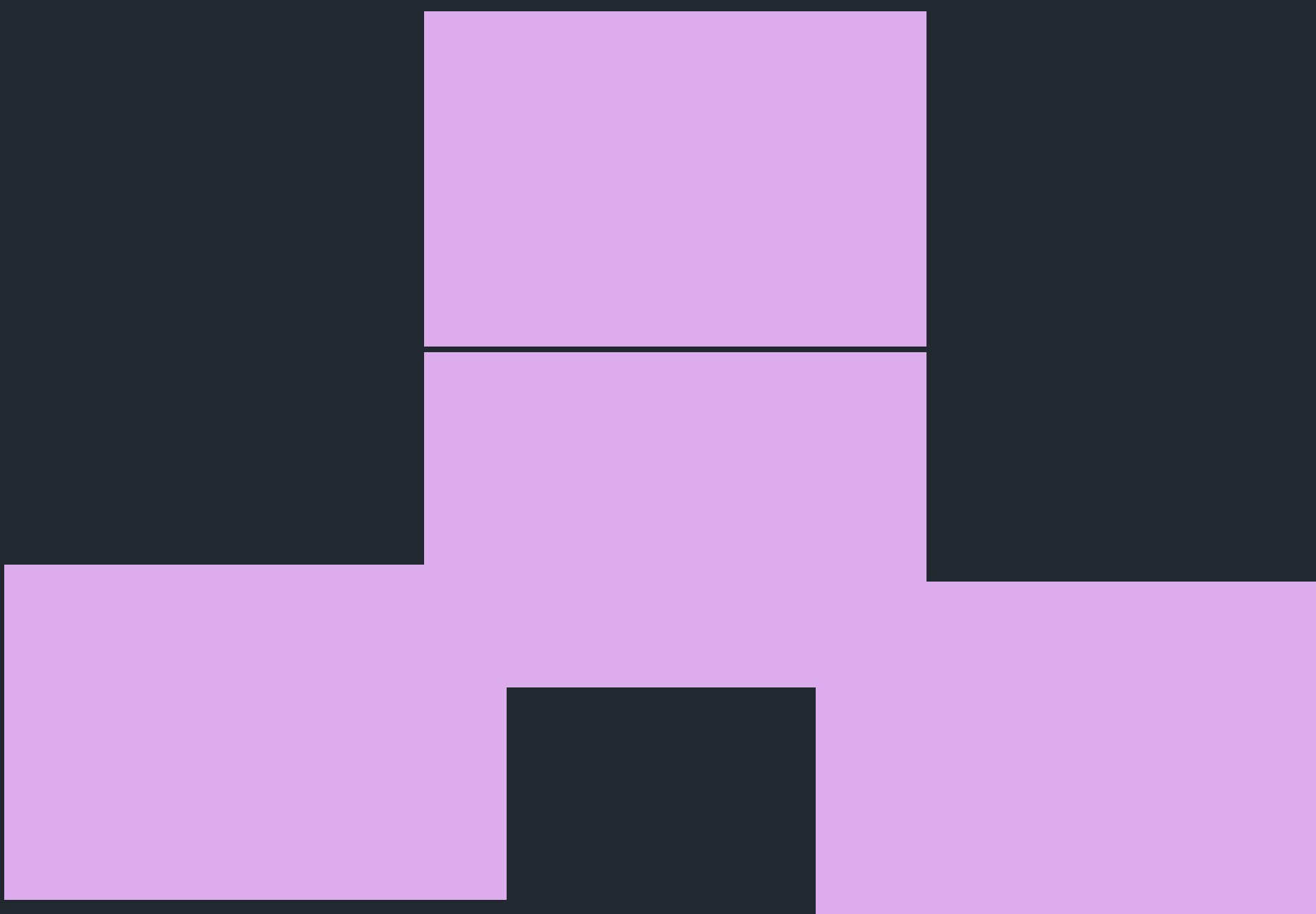
Once the **reference-free eval system** works,   
it can become part of the pipeline

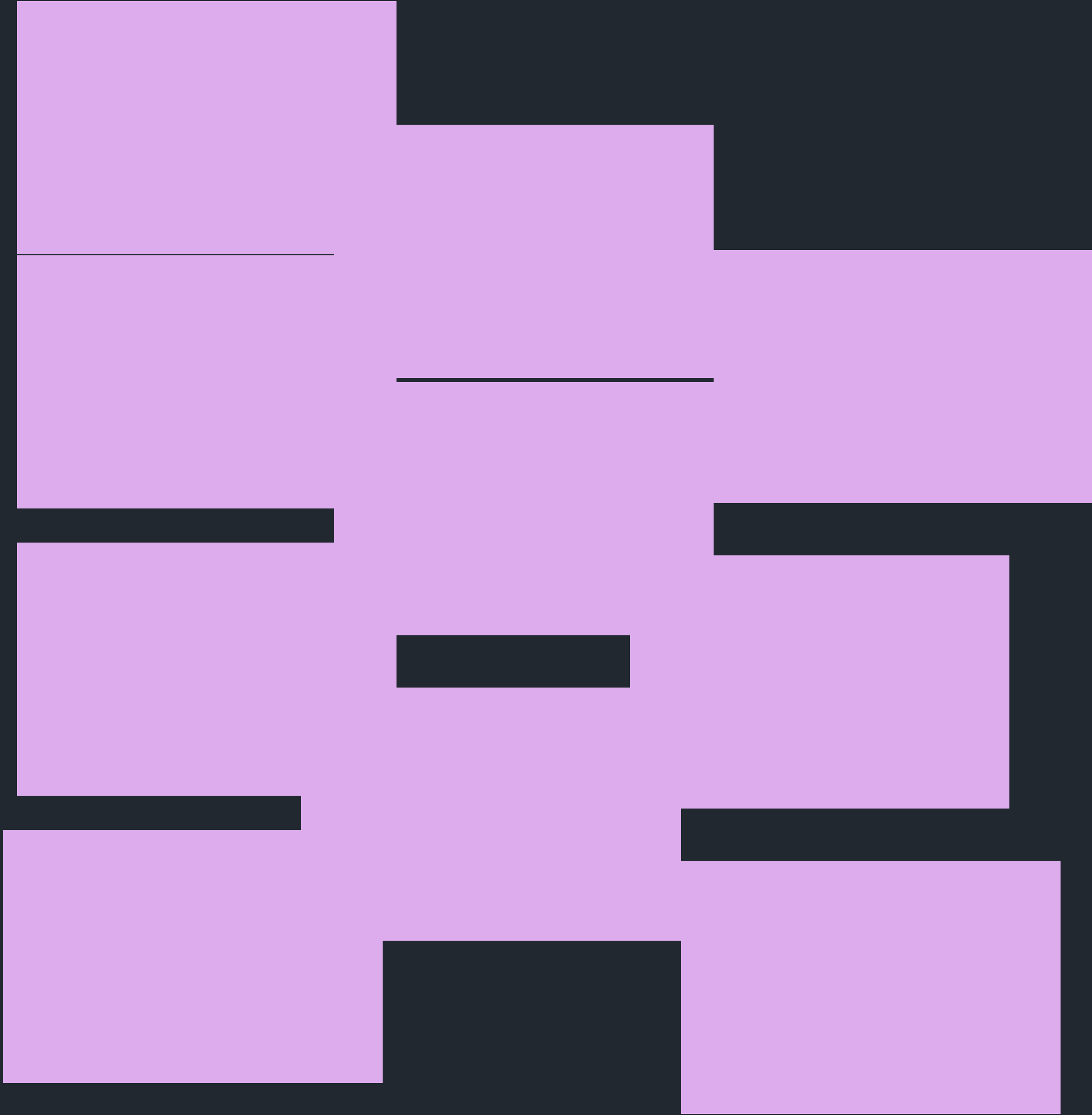


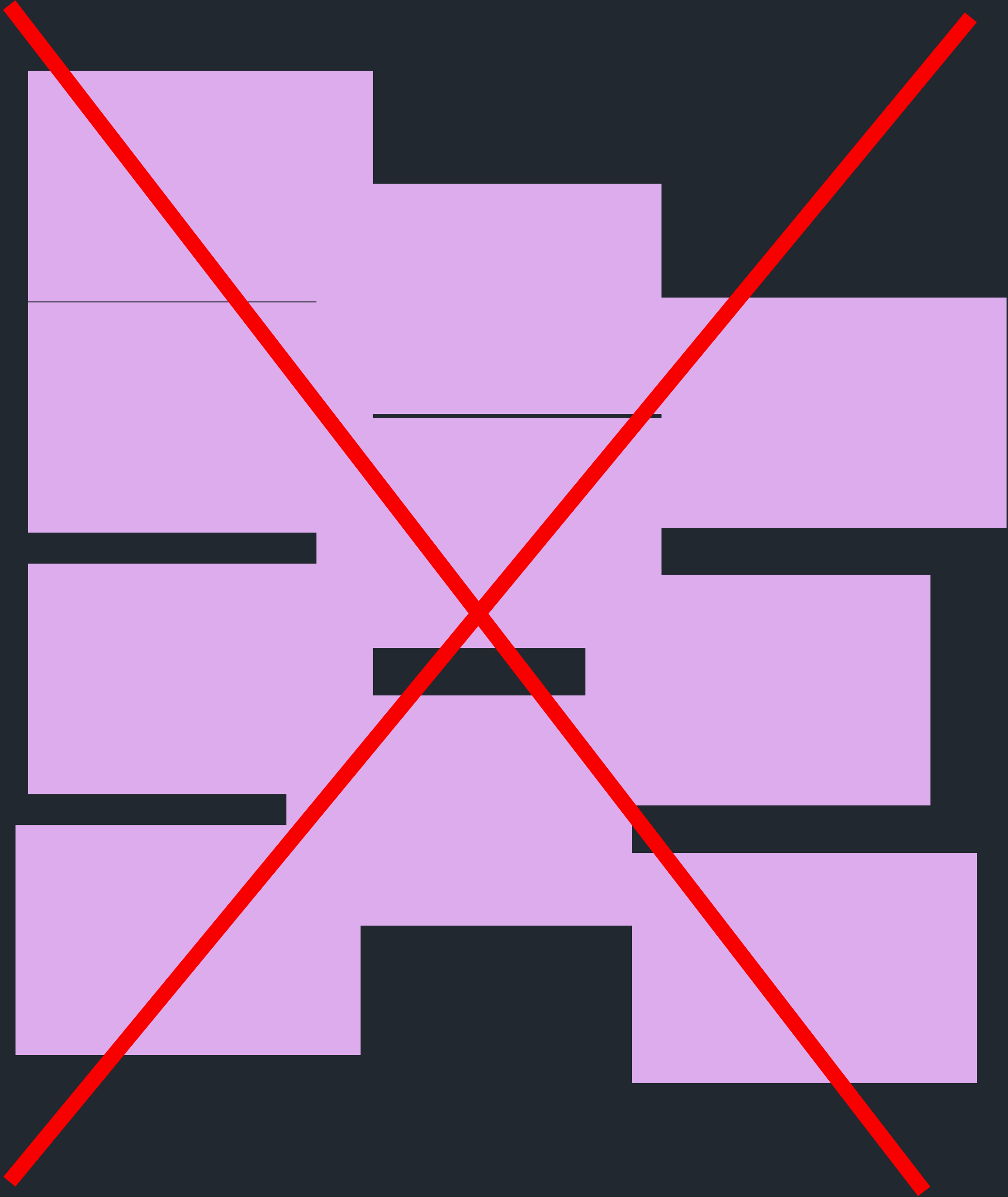


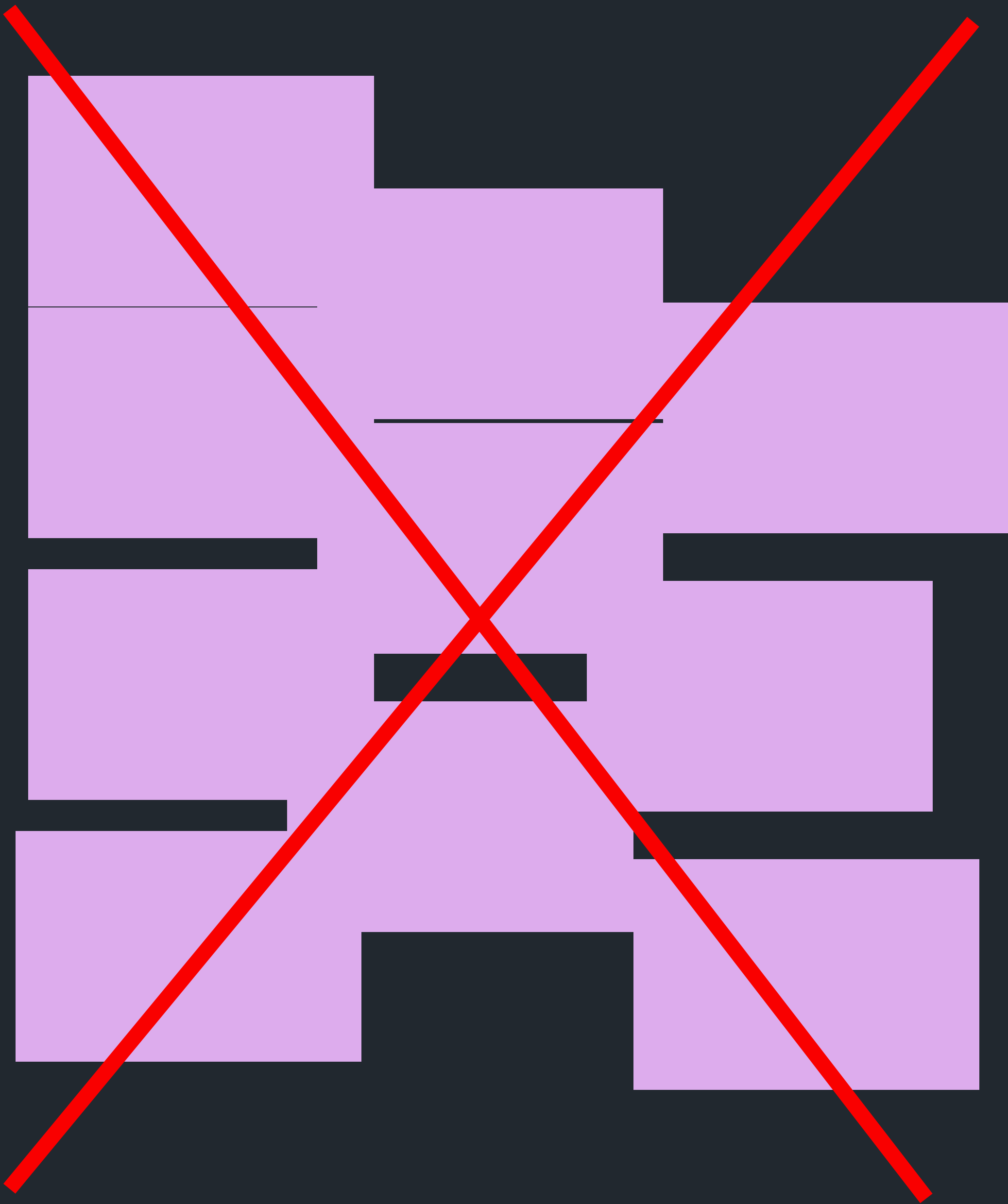
# The impact



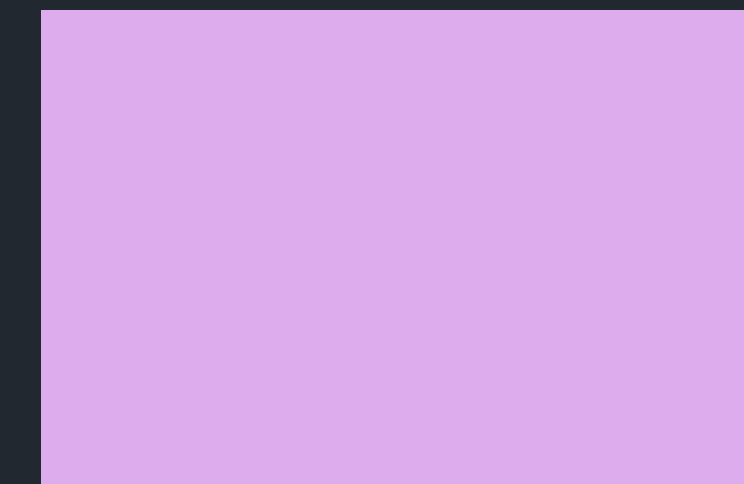








review 10,000+ outputs



< 10 clinical staff



94%

Alignment

*between AI and human reviews*



94%

Alignment

*between AI and human reviews*

95%

Alignment

*between human and human reviewers*



## Q: Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?

The medical record shows a brain MRI from 17/5/2022 that demonstrates 'multiple foci of T2/FLAIR hyperintensity in the infratentorial, juxtacortical and periventricular white matter' which is noted to be 'consistent with multiple sclerosis.' This confirms prior brain MRI findings suspicious for MS.



## Q: Does the patient have a prior brain MRI suspicious for Multiple Sclerosis?

The medical record shows a brain MRI from 17/5/2022 that demonstrates 'multiple foci of T2/FLAIR hyperintensity in the infratentorial, juxtacortical and periventricular white matter' which is noted to be 'consistent with multiple sclerosis.' This confirms prior brain MRI findings suspicious for MS.



While there is a brain MRI from 17/5/2022 described as 'consistent with multiple sclerosis', this finding is documented in the context of established Multiple Sclerosis diagnosis. The medical record confirms a definitive MS diagnosis, precluding classification of these imaging findings as 'suspicious for' MS. Rather, these MRI findings represent characteristic radiological features consistent with the patient's known MS diagnosis.





F1-Score

95.73%

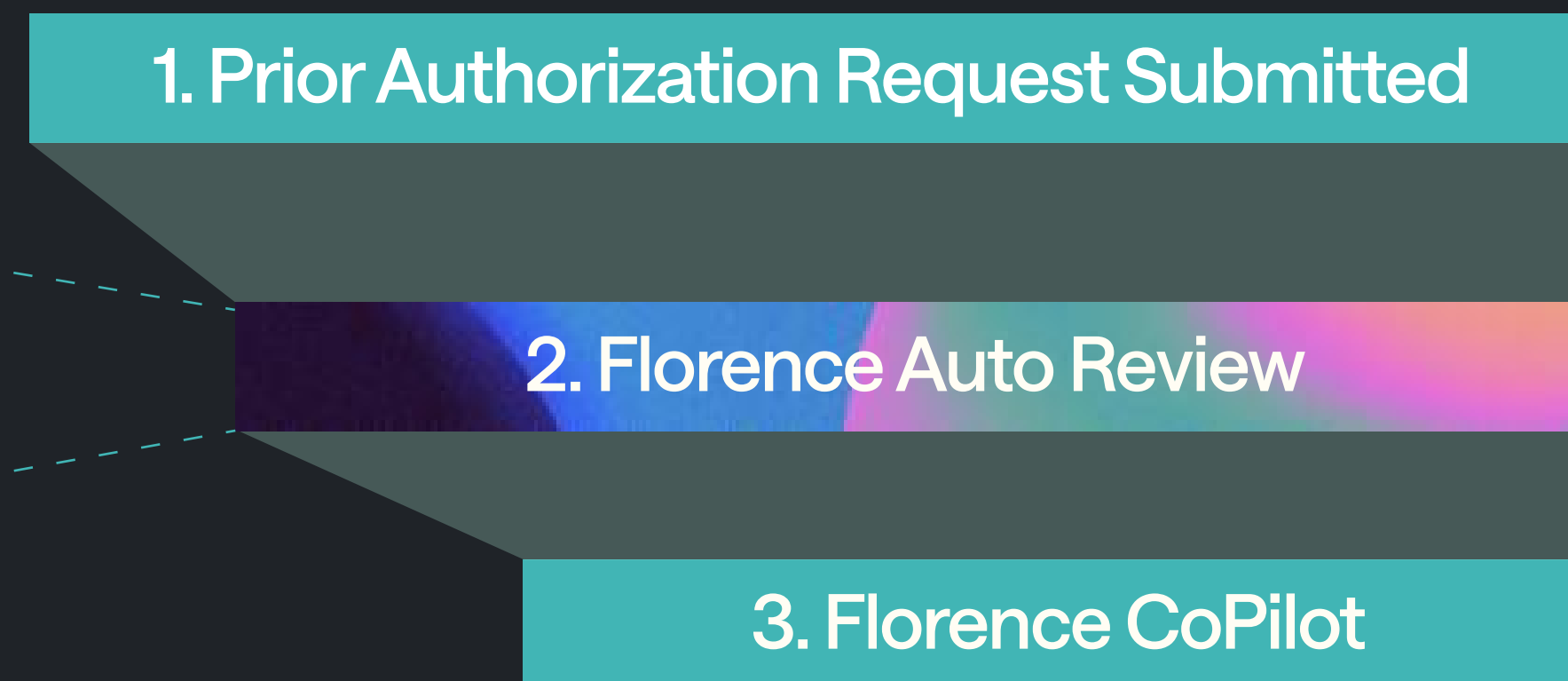
*for automated approvals across outpatient reviews*

Precision

96.46%

Recall

95.01%



High-confidence approvals are automated by Florence, deflecting the need for human review, getting to an approval faster

Florence escalates potential denials and cases with low certainty for additional human review

*Statistics across a random sample of 872 prior authorization reviews - Florence deployed in a payer organization that reviews 6M prior authorizations/year.*



**And our customers trust our product**



And our customers ~~trust~~ our product



# Florence delivers 5x time saving in MedWatch case study

Time Saving

77%

*across outpatient reviews*

Accuracy

97%

nurse agreement

Happiness

92%

nurse satisfaction

*When the nurses were told that they could carry on using Florence after the time study...*

“Thank God. We’re the lucky ones”.



## Build a system

Don't just report  
performance

## Evaluate on live production data

Proactive, not  
reactive

## Get the best reviewers and empower them

Build or buy  
bespoke tooling



**MVP -> serving customers at scale**



Thank you

[chris@anterior.com](mailto:chris@anterior.com)

[Anterior.com/Company](https://anterior.com/Company)

