



When LLMs Meet Mental Health Measurement

Adam Kolář

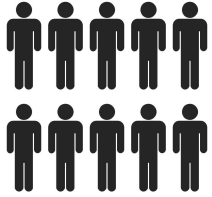


Outline

- Intro to HAM-A & clinical trials
- Lessons learned
- What the real issues in clinical practice are

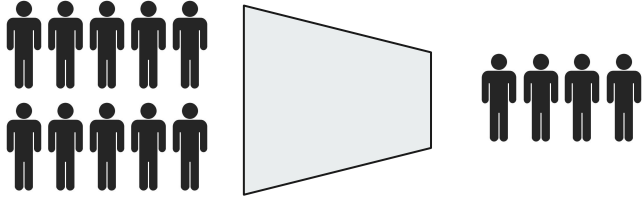
Clinical trials

Is the medication working?



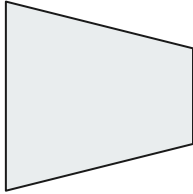
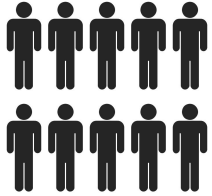
Is the medication working?

Inclusion/exclusion
criteria



Is the medication working?

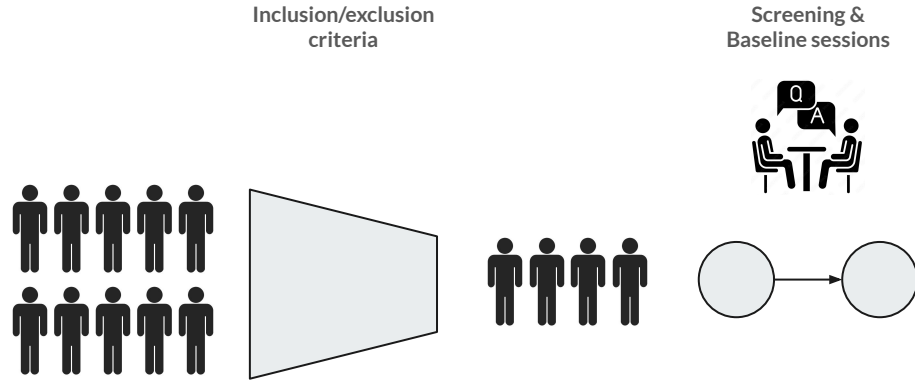
Inclusion/exclusion
criteria



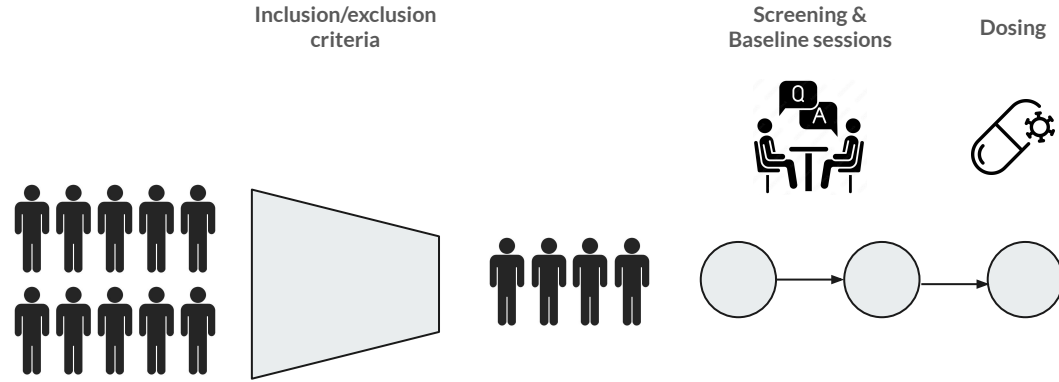
HAM-A score

- Minimal: 0 – 7
- Mild: 8 – 14
- **Moderate 15 – 23**
- Severe: ≥ 24 severe (up to 56)

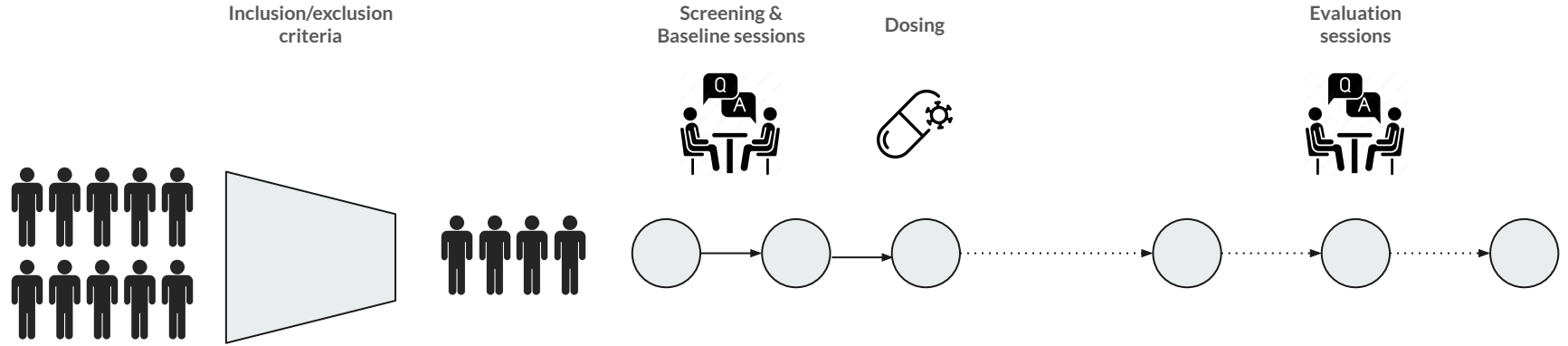
Is the medication working?



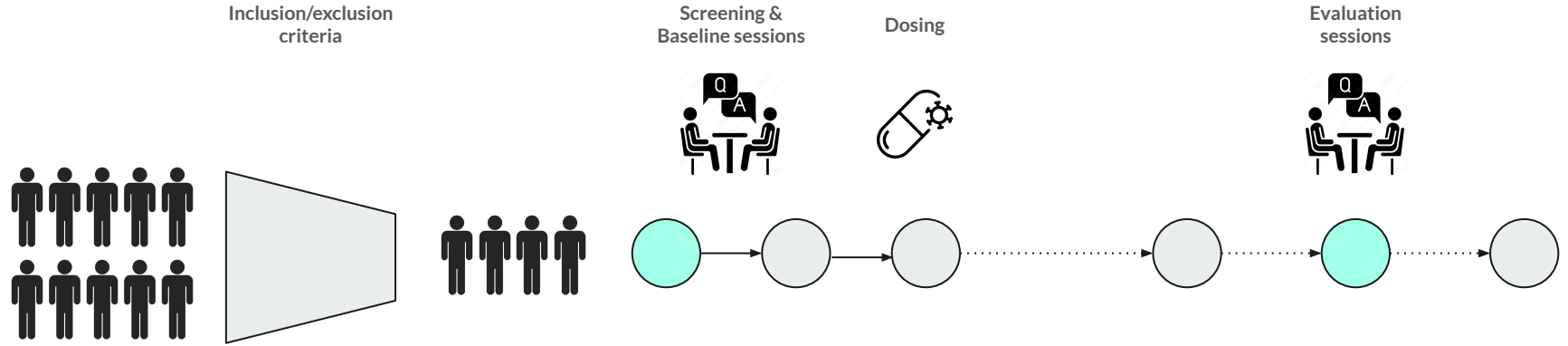
Is the medication working?



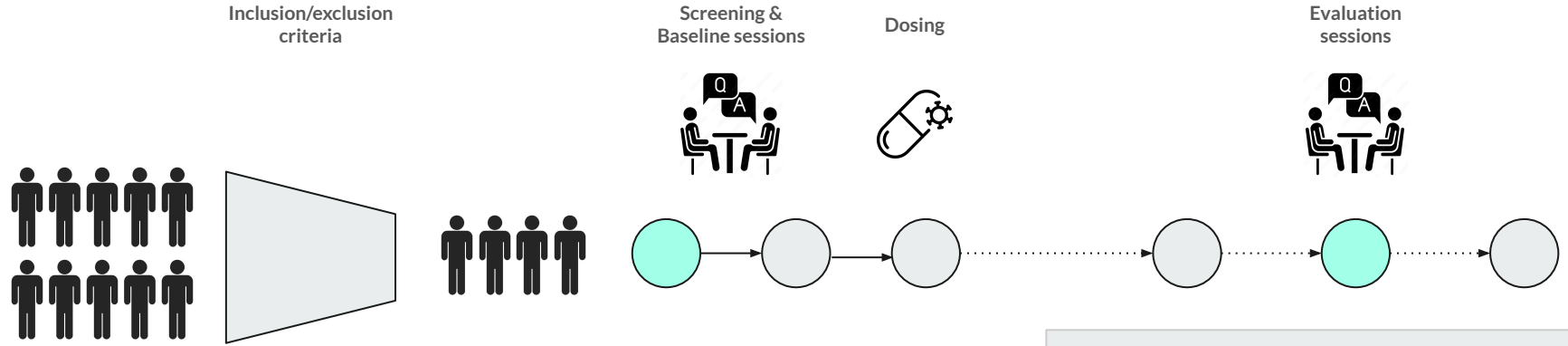
Is the medication working?



Is the medication working?



Is the medication working?



Treatment response

- Non responder: <25% reduction
- Partial responder: >25% and <50%,
- Responder and non-remitter: >50% and score >8
- Remitter: score <7

HAM-A

**What is the challenge?
And why it matters?**

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

STRUCTURED INTERVIEW GUIDE FOR THE HAMILTON ANXIETY SCALE (SHAS)

James B. H. Williams, D. M.D.

INTERVIEWER: The questions for each item that appear in bold type should be asked exactly as written. Follow-up questions are provided for further exploration of symptoms. For each response endorsed, use the additional probes at the top of each page to determine the frequency and severity of the symptoms, including how long it takes, how much distress it causes, and if the symptoms have occurred any previous time. These questions should be asked only if you have enough information to rate the item confidently. You may choose to skip any item following questions to obtain necessary information. For each of the 16 items, you may choose to skip any item following questions to obtain necessary information. You do not need to repeat questions about these symptoms unless you have additional information to use them severely.

All of the items have the same anchor points. The following may be useful as a guide to rating item severity.

ABSENT	Score 0 symptoms are absent, insignificant, or clearly not a concern at all.
MILD	Score 1 symptoms are infrequent, with no impairment and no more than mild distress.
MODERATE	Score 2 symptoms are more frequent, with moderate distress or impairment.
SEVERE	Score 3 symptoms are severe and persistent or result in severe distress or marked impairment in functioning.
VERY SEVERE	Score 4 symptoms are incapacitating.

For each symptom, ask:

- Tell me what that was like. Can you give me some examples?
- How bad has this been over the past week?
- How much has it bothered you? Has it caused you any problems?
- How much has it bothered you? Has it caused you any problems?
- How much has it bothered you or others have you had to take care of the past week?

NOTES:

Time period: Ratings should be based on the patient's condition during the past week. Unlike the HAM-D, changes from one week to the next should not be rated positively on the HAM-A. However, symptoms should not be rated negatively if they are clearly due to a disease unrelated to anxiety. A. Hypothetical symptoms due to physical illness.

Specific attacks: If the patient has panic attacks, this could affect the ratings of many of the symptoms. It is recommended that you consider the hour amount of time during the past week that the panic attacks symptoms occurred, as well as their severity. For example, a patient who has a few severe but short-acted panic attacks during the week, but who otherwise does not have many anxiety symptoms, would probably not have a very high total HAM-A score.

© 1997 by University of Pittsburgh Medical Center. All rights reserved. This instrument is a copyrighted work of the University of Pittsburgh Medical Center. It is not to be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without the prior written permission of the University of Pittsburgh Medical Center. For further information, contact: Health Services, Room 3500, University of Pittsburgh Medical Center, Pittsburgh, PA 15261.

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

OVERVIEW: For use with sub-symptom questions about the past week. How have you been feeling about this item (DASH OF MEASURE)? IF OUTPATIENT: How have you been feeling? IF NOT: Why not?

Suggested follow-up questions:

- Tell me what that was like. Can you give me some examples?
- How bad has this been over the past week?
- How much has it bothered you? Has it caused you any problems?
- How much of the time or how often have you had this the past week?

ABSENT	Score 0 symptoms are absent, insignificant, or clearly not a concern at all.
MILD	Score 1 symptoms are infrequent, with no impairment and no more than mild distress.
MODERATE	Score 2 symptoms are more frequent, with moderate distress or impairment.
SEVERE	Score 3 symptoms are severe and persistent or result in severe distress or marked impairment in functioning.
VERY SEVERE	Score 4 symptoms are incapacitating.

ANXIETY MOOD
Irritable, irascible, or easily annoyed; hostile; suspicious; irritable.

1 = not present
2 = mild
3 = moderate
4 = very severe

IN THE PAST WEEK, HOW MUCH HAVE YOU BEEN WORRYING?
(What have you been worried about? How bad has it been to sleep worrying?)

IF UNKNOWN: How much have you been afraid that the world is going to happen?

How have you been feeling nervous or restless this past week?

How have you been feeling itchy this past week?

IF SCORED 1-4 ABOVE, FOR CONVEY, ASK: How long have you been feeling this way?

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

IN THE PAST WEEK, HOW MUCH HAVE YOU FELT TIRED, HEAVY OR, OR SLEPT?

How often do you get tired easily?

How much have you been bothered by any of these things?

- waking easily
- breathing?
- feeling restless?
- not being able to relax?

IN THE PAST WEEK, HOW HAVE YOU BEEN ABLE TO ...

- do the work?
- get along?
- do things you like to do?
- do things you have to do?
- do things you don't want to do?

How have you had any other specific fears this past week?

NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

Now let's talk about your sleep.

In the past week, how many times have you had trouble falling asleep?

How long has it been taking you to fall asleep?

IF SCORED THREE OR HIGHER: How many nights this past week did this happen?

In the past week, how many times have you been waking up in the middle of the night? IF YES: How long did you stay awake?

IF UNKNOWN: In the past week, how many times have you been waking up in the middle of the night? IF YES: How long did you stay awake?

How many times have you been waking up in the middle of the night? IF YES: How long did you stay awake?

How about having bad dreams or nightmares?

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

IN THE PAST WEEK, HAVE YOU HAD TO URGENTLY USE THE BATHROOM? IF YES: How often have you had to do this?

How has your interest in sex been this past week? Do not think about performance, but about your interest in sex.

IS THIS A CHANGE FOR YOU? IF YES: How much of a change?

FOR MEN: When you have sex, do you have less or more of the following? IF YES: How much more or less?

FOR MEN: Sometimes when you have sex, do you have less or more of the following? IF YES: How much more or less?

FOR MEN: Sometimes when you have sex, do you have less or more of the following? IF YES: How much more or less?

FOR MEN: Sometimes when you have sex, do you have less or more of the following? IF YES: How much more or less?

FOR MEN: Sometimes when you have sex, do you have less or more of the following? IF YES: How much more or less?

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

IN THE PAST WEEK, HAVE YOU HAD TROUBLE CONCENTRATING?

IF YES: Can you give me some examples?

How about trouble reading - like a book or a newspaper? Do you need to read things over and over again?

How have you had any trouble following a conversation?

How have you had trouble remembering things this past week?

IF UNKNOWN: What about remembering appointments or events you have to do?

INTELLECTUAL
(Difficulty concentrating, poor memory)

0 = not present
1 = mild
2 = moderate
3 = severe
4 = very severe

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

IN THE PAST WEEK, HAVE YOU FELT DEPRESSED, OR SAD?

Can you describe the feeling? How bad has it been?

IF YES: Does the feeling fit something good happen?

How have you been feeling about the future?

How have you been feeling discouraged or pessimistic?

What have your thoughts been?

How have you been crying this past week?

How have you been interested in things, or not enjoying things you usually enjoy things?

How have you been these this past week when you have had something other than worry? (Is it with or without any other symptoms?)

This past week, how have you been feeling better or worse about things you usually enjoy? IF UNKNOWN: How much worse? How many days has this been the pattern?

DEPRESSED MOOD
Loss of interest in activities, hobbies, depression, early waking, dull affect.

0 = not present
1 = mild
2 = moderate
3 = severe
4 = very severe

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

IN THE PAST WEEK, ...

- your heart beat, skipped or pounding?
- How have you had pain in your chest?
- Had any throbbing blood vessels?
- Any lightheaded feelings?

IN THE PAST WEEK, HOW HAVE YOU BEEN ...

- pressure or tightness in your chest?
- churning feelings?
- What about tightness?
- How have you had shortness of breath?

CARDIOVASCULAR SYMPTOMS
Chest pain, palpitations, pain in chest, throbbing vessels, lightheadedness.

0 = not present
1 = mild
2 = moderate
3 = severe
4 = very severe

RESPIRATORY SYMPTOMS
Pressure or constriction in chest, churning feelings, tightness, oppression.

0 = not present
1 = mild
2 = moderate
3 = severe
4 = very severe

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

IN THE PAST WEEK, HAVE YOU HAD ...

- trouble swallowing?
- stomach pain or fullness?
- heart?
- nausea?
- acid reflux?
- burning or itching in your stomach?
- loose bowels?
- constipation?

How have you been these this past week when you have had something other than worry? (Is it with or without any other symptoms?)

GASTROINTESTINAL SYMPTOMS
Difficulty in swallowing, acid, abdominal pain, burning sensation, obstructive feelings, nausea, vomiting, bloating, loss of appetite, loss of weight, constipation.

0 = not present
1 = mild
2 = moderate
3 = severe
4 = very severe

NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

PROCTOR MEASURES Date 11-06 Meeting ID 7848 Center Name WHS

RATING BASED ON OBSERVATION (CIRCLE INTERVIEW (CIRCLE SYMPTOM PRESENT):

FOR TELEPHONE ASSESSMENT:
During the interview, have you been sleeping at all, or having trouble falling asleep? How have you been doing everything else during the past week or so, like work, school, or other things you have to do?

IF YES: How bad has that been?

How have your hands been shaky?

What about sweating or feeling the need to urinate?

If you looked in a mirror right now, would your face look worse? IF NO: Would it be noticed or hard?

Do you think you look pale?

14. INQUIRY AT INTERVIEW
(Inquiry, restlessness or pacing, motor or hand tremor, face, neck, chest, back, legs or feet, sweating, shaking, skin broken, skin, closed pupils, incontinence, etc.)

0 = not present
1 = mild
2 = moderate
3 = severe
4 = very severe

MEASUREMENT ID: SHAS_UH_English_U1_2002_Ang_10

TIME ENDED SHAS: 5:58 AM PM OT FT

TOTAL HAMILTON ANXIETY SCORE: 37



STRUCTURED INTERVIEW GUIDE FOR THE HAMILTON ANXIETY SCALE (SIGH-A)

Janet B.W. Williams, D.S.W.

INTERVIEWER: The questions for each item that appear in bold type should be asked exactly as written. Follow-up questions are provided for further exploration of symptoms. For each symptom endorsed, use the additional probes at the top of each page to determine the frequency and severity of the symptom, including how bad it's been, how much distress it has caused, and if the symptom has caused any impairment. These questions should be asked until you have enough information to rate the item confidently. You may also have to add your own follow-up questions to obtain necessary information. For some of the HAM-A items, you may find you have already asked about some of the symptoms (for a previous item). You do not need to repeat questions about these symptoms unless you need additional information to rate their severity.

All of the items have the same anchor points. The following may be useful as a guide to rating item severity:

ABSENT	Score 0 if symptoms are absent, insignificant, or clearly due to causes other than anxiety
MILD	Score 1 if symptom is infrequent, with no impairment and no more than mild distress
MODERATE	Score 2 if symptom is more frequent, with moderate distress or limited interference with usual activities
SEVERE	Score 3 if symptom is severe and persistent or results in severe distress or marked impairment in functioning
VERY SEVERE	Score 4 if symptom is incapacitating

For each symptom, ask:

- Tell me what that was like. Can you give me some examples?
- How bad has this been over the past week?
- How much has it bothered you? Has it caused you any problems?
- How much of the time or how often have you had this over the past week?

NOTES:

Time period. The ratings should be based on the patient's condition during the past week. Unlike the HAM-D, change from usual self is not required for most items to be rated positively on the HAMA. However, symptoms should not be rated positively if they are clearly due to a cause unrelated to anxiety, e.g., respiratory symptoms due to pneumonia.

Panic attacks. If the patient has panic attacks, this could affect the ratings of many of the symptoms. It is recommended that you consider the total amount of time during the past week that the panic attack symptoms occurred, as well as their severity. For example, a patient who has a few severe but short-lived panic attacks during the week, but who otherwise does not have many anxiety symptoms, would probably not have a very high total HAM-A score.

This instrument provides an interview guide for the Hamilton Anxiety Scale (Hamilton M: The assessment of anxiety states by rating. *Br J Med Psychol* 32:50-55, 1959). Hamilton M: The diagnosis and rating of anxiety. In *Studies of Anxiety*. MH Ludes, Ed. Hoadley Bros., Kent, 1969). The anchor point descriptions for the scale have been taken from the ECDEU Assessment Manual (Guy, William, *ECDEU Assessment Manual for Psychopharmacology*. Revised 1976, DHEW Publication No. (ADM) 76-338), except that "agiting" and "dyspnea", which appear twice in that version, have been taken out of the item "cardiorespiratory symptoms," and left under "respiratory symptoms." It has been demonstrated that an interview guide strengthens the reliability of individual scale items of the Ham-D (Williams JBW: A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatry* 45:742-747, 1988). This work was supported in part by Biomedical Research Support Grant #903-E7595 and NIMH Grant #1 P50 MH 42520. Kenneth A. Kobak, PhD and Joshua D. Lipsitz, PhD contributed to revisions of this interview guide. For further information, contact Mapi Research Trust (Internet: <https://eguide.mapi-trust.org>).



**STRUCTURED INTERVIEW GUIDE FOR THE HAMILTON ANXIETY SCALE
(SIGH-A)**

Janet B.W. Williams, D.S.W.

INTERVIEWER: The questions for each item that appear in bold type should be asked exactly as written. Follow-up questions are provided for further exploration of symptoms. For each symptom endorsed, use the additional probes at the top of each page to determine the frequency and severity of the symptom, including how bad it's been, how much distress it has caused, and if the symptom has caused any impairment. These questions should be asked until you have enough information to rate the item confidently. You may also have to add your own follow-up questions to obtain necessary information. Examples of the HAMA items you may find useful have already been selected about some of the

ABSENT	Score 0 if symptoms are absent, insignificant, or clearly due to causes other than anxiety
MILD	Score 1 if symptom is infrequent, with no impairment and no more than mild distress
MODERATE	Score 2 if symptom is more frequent, with moderate distress or limited interference with usual activities
SEVERE	Score 3 if symptom is severe and persistent or results in severe distress or marked impairment in functioning
VERY SEVERE	Score 4 if symptom is incapacitating

NOTES:

Time period. The ratings should be based on the patient's condition during the past week. Unlike the HAM-D, change from usual self is not required for most items to be rated positively on the HAMA. However, symptoms should not be rated positively if they are clearly due to a cause unrelated to anxiety, e.g., respiratory symptoms due to pneumonia.

Panic attacks. If the patient has panic attacks, this could affect the ratings of many of the symptoms. It is recommended that you consider the total amount of time during the past week that the panic attack symptoms occurred, as well as their severity. For example, a patient who has a few severe but short-lived panic attacks during the week, but who otherwise does not have many anxiety symptoms, would probably not have a very high total HAMA score.

This instrument provides an interview guide for the Hamilton Anxiety Scale (Hamilton M: The assessment of anxiety states by rating, *Br J Med Psychol* 32:50-55, 1959; Hamilton M: The diagnosis and rating of anxiety, in *Studies of Anxiety*, MH Ludes, Ed., Hurdley Bros., Kent, 1969). The anchor point descriptions for the scale have been taken from the ECDEU Assessment Manual (Guy, William, *ECDEU Assessment Manual for Psychopharmacology*, Revised 1976, DHEW Publication No. (ADM) 76-338), except that "agiting" and "dyspnea," which appear twice in that version, have been taken out of the item "cardiorespiratory symptoms," and left under "respiratory symptoms." It has been demonstrated that an interview guide strengthens the reliability of individual scale items of the Ham-D (Williams JBW: A structured interview guide for the Hamilton Depression Rating Scale, *Arch Gen Psychiatry* 45:742-747, 1988). This work was supported in part by Biomedical Research Support Grant #903-E7595 and NIMH Grant #1 P50 MH 42520. Kenneth A. Kobak, PhD and Joshua D. Lipsitz, PhD contributed to revisions of this interview guide. For further information, contact Mapi Research Trust (Internet: <https://esrovide.mapi-trust.org>).



<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ severe 4 - very severe</p>



<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
--	---

<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ severe 4 - very severe</p>
---	---



<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
--	---

<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ severe 4 - very severe</p>
---	---



<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ severe 4 - very severe</p>



Replace slow and expensive double check with fast and precise scoring (and other features)

<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe 4 - very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe 4 - very severe</p>



- Did you referenced past week?
- Do you ask the questions precisely?

<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ severe 4 - very severe</p>



- Did you referenced past week?
- Do you ask the questions precisely?
- Do you ask all the questions?

<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ severe 4 - very severe</p>



- Did you referenced past week?
- Do you ask the questions precisely?
- Do you ask all the questions?
- How consistent and precise the scoring is?

<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe 4 - very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe 4 - very severe</p>

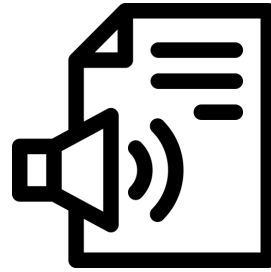


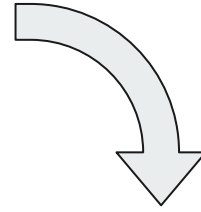
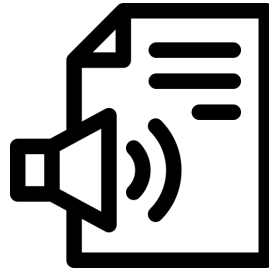
- Did you referenced past week?
- Do you ask the questions precisely?
- Do you ask all the questions?
- How consistent and precise the scoring is?

Evaluations need double check and fast feedback loop

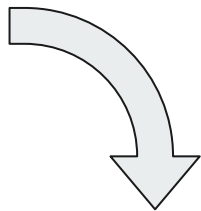
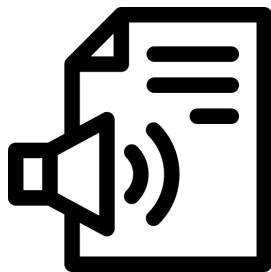
<p>In the past week, how much have you felt tense, keyed up, or on edge?</p> <p>Have you gotten tired easily?</p> <p>How much have you been bothered by any of these things:</p> <ul style="list-style-type: none">• being startled easily?• crying easily?• trembling?• feeling restless?• not being able to relax?	<p>2. TENSION (feelings of tension, fatigability, startle response, moved to tears easily, trembling, feelings of restlessness, inability to relax):</p> <p>0 - not present 1 - mild 2 - moderate 3 - severe ④ very severe</p>
<p>This past week, have you been afraid of . . .</p> <ul style="list-style-type: none">• the dark?• of strangers?• of being left alone?• of animals?• of traffic?• of crowds? <p>Have you had any other specific fears this past week?</p> <p>NOTE: INCLUDE ANY IRRATIONAL ANXIETY ABOUT OBJECTS OR SITUATIONS.</p>	<p>6. FEARS (of dark, of strangers, of being left alone, of animals, of traffic, of crowds):</p> <p>0 - not present 1 - mild 2 - moderate ③ - severe 4 - very severe</p>

Could LLM help?



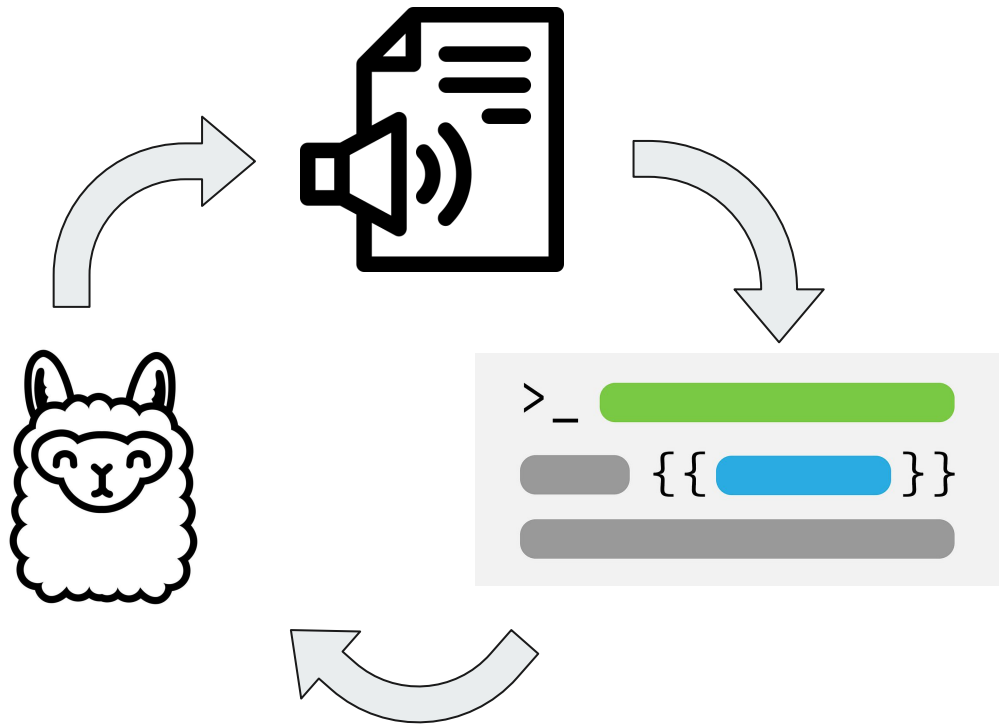


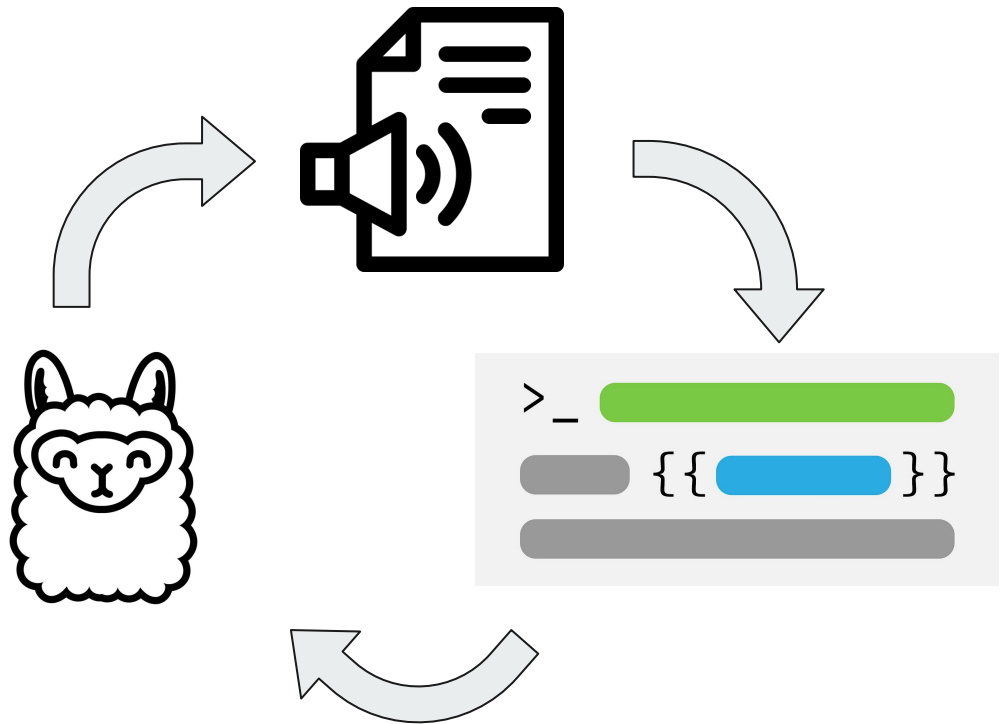
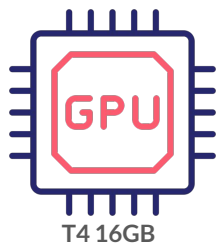
```
> _ [green bar]  
[gray bar] { [blue bar] }
```



```
> _ [REDACTED]  
[REDACTED] { { [REDACTED] } }  
[REDACTED]
```





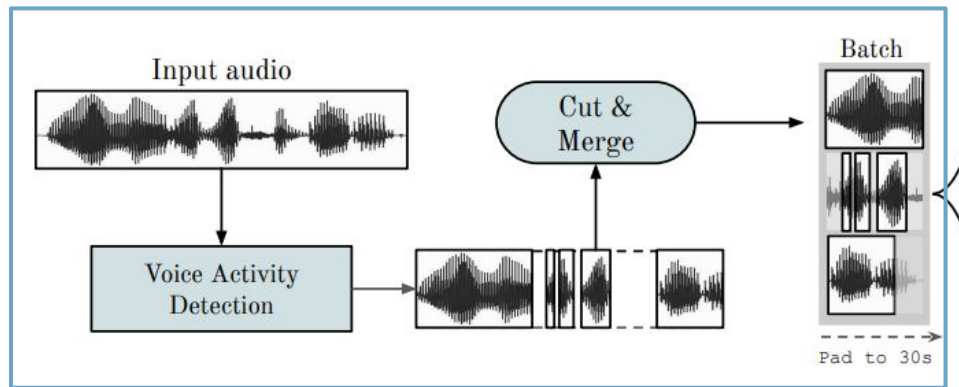


Audio Processing

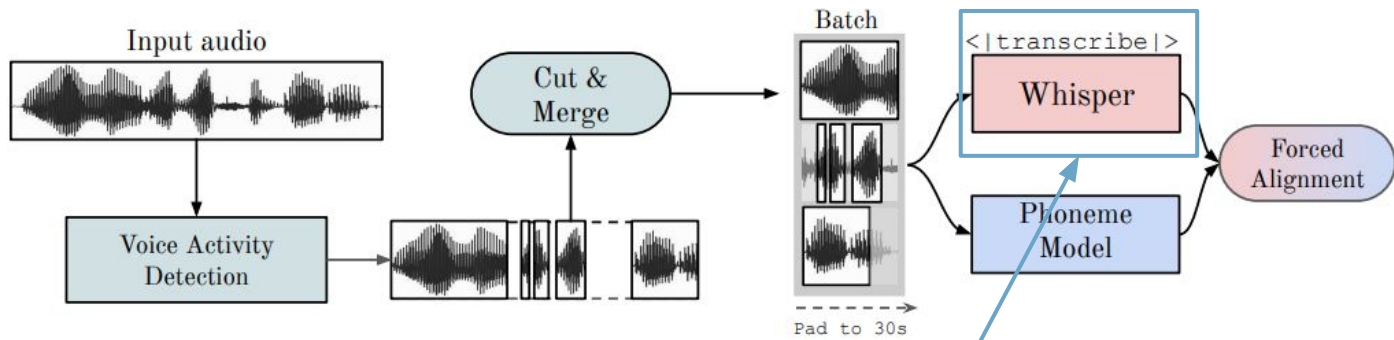


Reasons for WhisperX

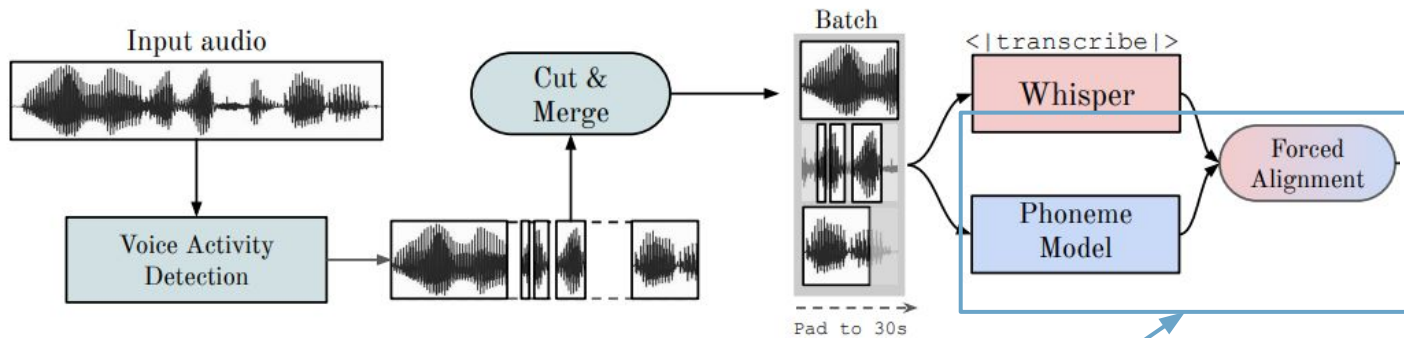
- Quality of transcription
- Speed
- Timestamps



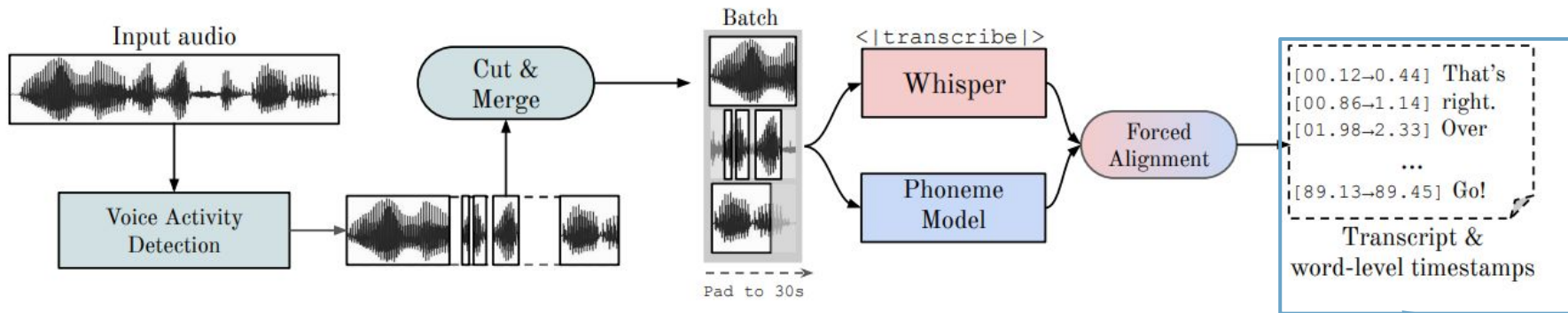
- VAD driven chunking



- VAD driven chunking
- Localized tokenizer
- No conditioning of previous text



- VAD driven chunking
- Localized tokenizer
- No conditioning of previous text
- Timestamp alignment using Wav2Vec



- VAD driven chunking
- Localized tokenizer
- No conditioning of previous text
- Timestamp alignment using Wav2Vec
- Hallucination detection

Prompting Dead Ends



Summarizing transcription first



Summarizing transcription first

- Iterative context aggregation
- No guiding principles on what is important -> missing details
- Prioritize more recent context
- Tries to make its own interpretation (e.g. urination is connected to insomnia)



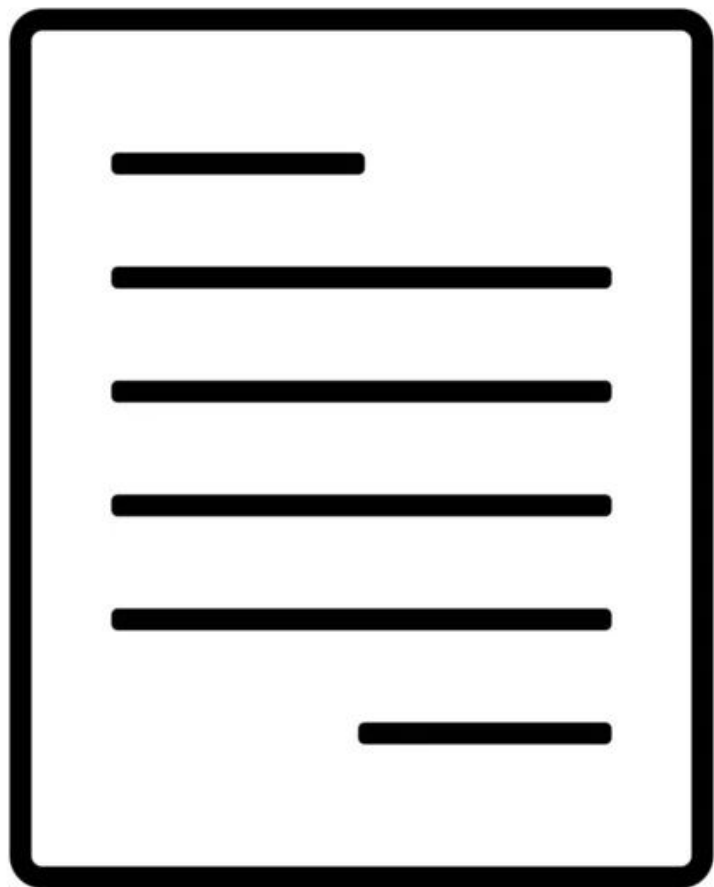
Zeroshot prediction of all the symptom



Zeroshot prediction of all the symptom

- Prompt is too big for fine tuning
- Hard to force model to cover all the symptoms
- Hallucinating new ones

Prompting strategy

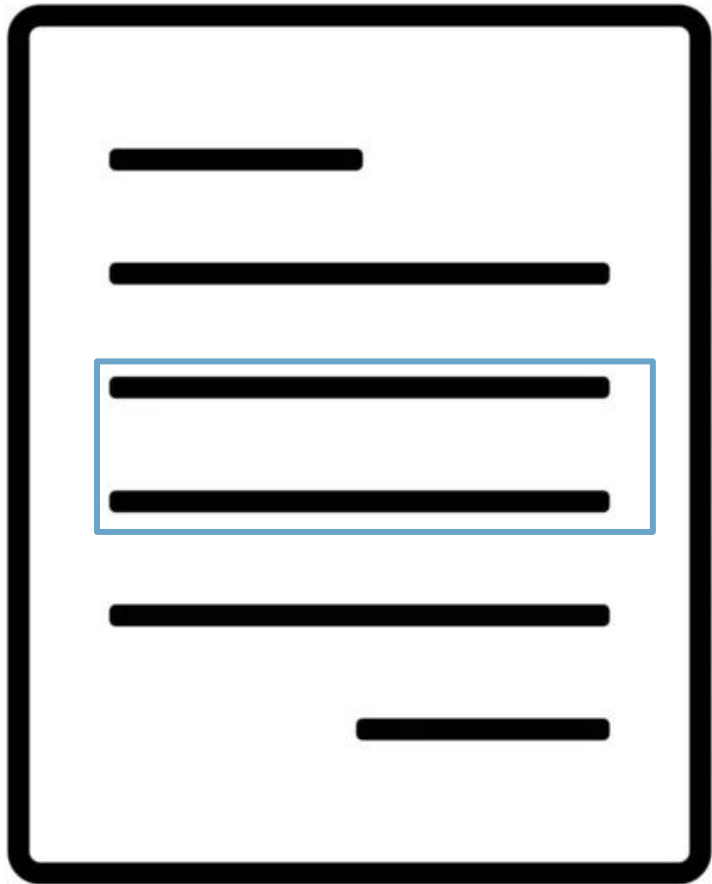


LLM

03:27 Anxiety

08:14 Tension

15:52 Fears



08:14 Tension



Symptom conditional
prompting

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. ****Focus****: Identify and evaluate only the utterances related to *{symptom}*.

2. ****Keywords****:
{keywords_section}

3. ****Severity Scale****:
{severity_scale}

4. ****Examples for Calibration****: Use the following examples to guide your evaluation:
{real_examples}

5. ****Response Format****: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input
{transcription}

Detected *{symptom}* severity:
{{
 "{symptom}": ""

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. **Focus**: Identify and evaluate only the utterances related to *{symptom}*.

2. **Keywords**:
{keywords_section}

3. **Severity Scale**:
{severity_scale}

4. **Examples for Calibration**: Use the following examples to guide your evaluation:
{real_examples}

5. **Response Format**: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input

{transcription}

Detected *{symptom}* severity:

```
{{  
  "{symptom}": ""  
}}
```

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. **Focus**: Identify and evaluate only the utterances related to *{symptom}*.

2. **Keywords**:
{keywords_section}

3. **Severity Scale**:
{severity_scale}

4. **Examples for Calibration**: Use the following examples to guide your evaluation:
{real_examples}

5. **Response Format**: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input
{transcription}

Detected *{symptom}* severity:
{{
 "{symptom}": ""

Keywords for Gastrointestinal symptoms

- stomach pain
- difficulty swallowing
- wind
- abdominal pain
- fullness
- nausea, vomiting, borborygmi
- looseness of bowels
- loss of weight
- constipation

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. ****Focus****: Identify and evaluate only the utterances related to *{symptom}*.

2. ****Keywords****:
{keywords_section}

3. ****Severity Scale****:
{severity_scale}

4. ****Examples for Calibration****: Use the following examples to guide your evaluation:
{real_examples}

5. ****Response Format****: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input
{transcription}

Detected {symptom} severity:
{{
 "{symptom}": ""

General severity scale

- **ABSENCE:** No symptoms or clearly not anxiety-related.
- **MILD:** Infrequent symptoms, mild distress, no impairment.
- **MODERATE:** More frequent symptoms, moderate distress, limited interference with activities.
- **SEVERE:** Severe, persistent symptoms, significant distress or impairment.
- **VERY SEVERE:** Incapacitating symptoms.

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. **Focus**: Identify and evaluate only the utterances related to *{symptom}*.

2. **Keywords**:
{keywords_section}

3. **Severity Scale**:
{severity_scale}

4. **Examples for Calibration**: Use the following examples to guide your evaluation:
{real_examples}

5. **Response Format**: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input
{transcription}

Detected *{symptom}* severity:
{{
 "{symptom}": ""

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. ****Focus****: Identify and evaluate only the utterances related to *{symptom}*.

2. ****Keywords****:
{keywords_section}

3. ****Severity Scale****:
{severity_scale}

4. ****Examples for Calibration****: Use the following examples to guide your evaluation:
{real_examples}

5. ****Response Format****: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input
{transcription}

Detected *{symptom}* severity:

```
{{  
  "{symptom}": ""  
}}
```

You are provided with a transcribed HAM-A interview between a clinician and a patient. Each line starts with the start and end time in minutes and seconds (mm:ss), followed by the transcription of the audio. The transcription may contain errors, typos, and unclear speaker identification. Your task is to estimate the severity of the patient's *{symptom}* based on this transcription.

Instructions:

1. **Focus**: Identify and evaluate only the utterances related to *{symptom}*.
2. **Keywords**:
{keywords_section}
3. **Severity Scale**:
{severity_scale}
4. **Examples for Calibration**: Use the following examples to guide your evaluation:
{real_examples}

5. **Response Format**: Provide your response as a Python dictionary:

```
{{  
  "{symptom}": "severity level"  
}}
```

For example:

```
{{  
  "{symptom}": "MILD"  
}}
```

- Do not add or remove any keys from the dictionary.
- Do not include additional text or explanations.
- Input may contain content related to other symptoms. Ignore everything not related to *{symptom}*.

Input

{transcription}

Detected *{symptom}* severity:

```
{{  
  "{symptom}": ""  
}}
```

Aggregated real examples for Gastrointestinal symptoms

VERY SEVERE

Intense nausea and vomiting due to anxiety, causing significant distress and interference with daily activities.

SEVERE

Frequent trouble swallowing and vomiting. Persistent stomach pain and gas with moderate discomfort. Constipation resolves itself.

MODERATE

Moderate stomach pain, fullness, and gas occurring 2–3 times weekly, causing minor discomfort.

MILD

Minor gas without distress, and intentional weight loss. No other gastrointestinal symptoms.

ABSENCE


No gastrointestinal symptoms such as swallowing difficulties, stomach pain, gas, nausea, or vomiting.



Takeaways

- Small LLMs struggle to consume long transcribes / prompts
- Structure & known formats like markdown, json, python dict help
- Few-shot can not cover variability of situations

Finetuning



Llama 3.x

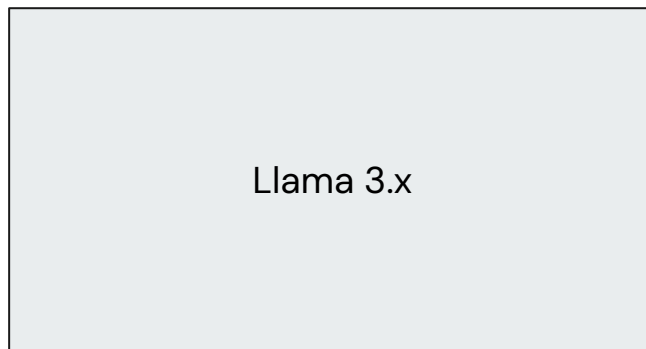
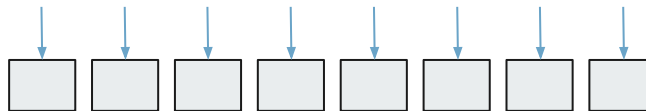
Llama 3.x



unslloth

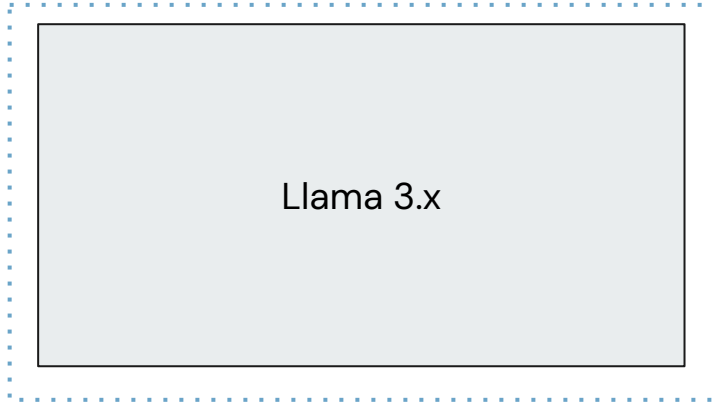
- 4NF QLoRA
- Kernels
- RoPE rescaling

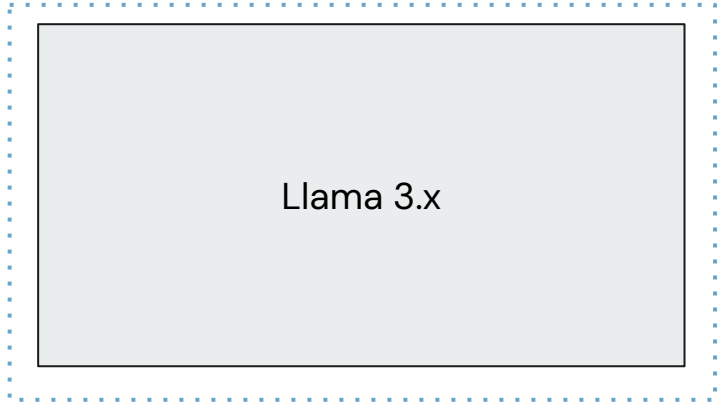
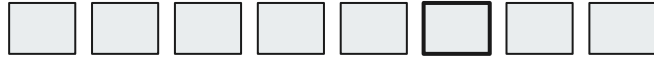
cross entropy



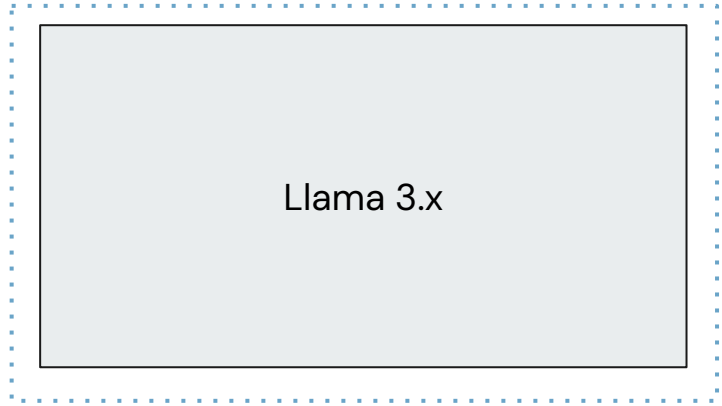
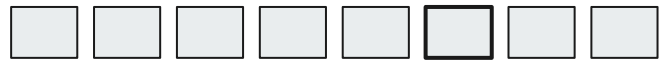
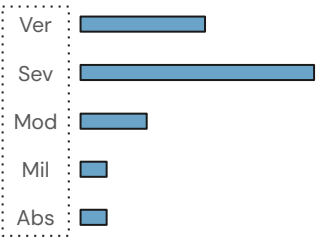
unslloth

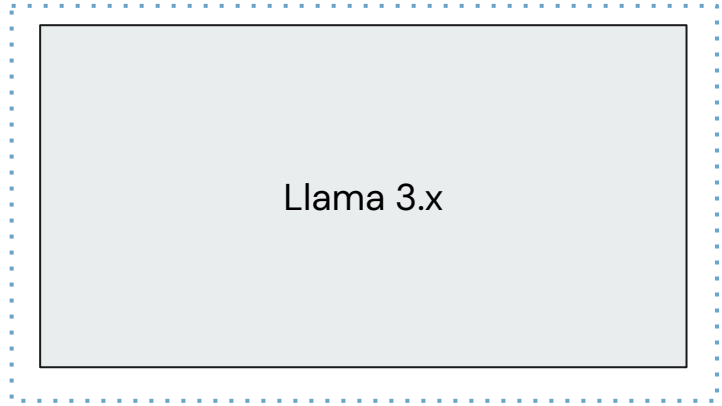
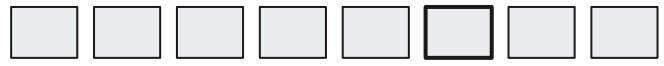
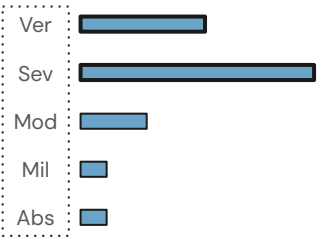
- 4NF QLoRA
- Kernels
- RoPE rescaling

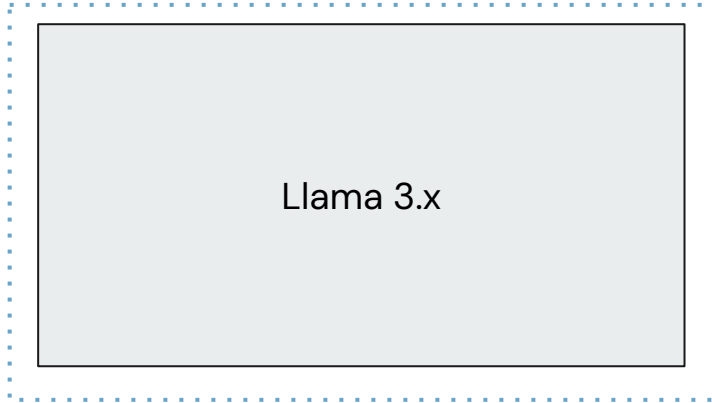
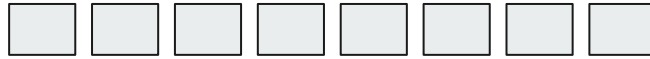




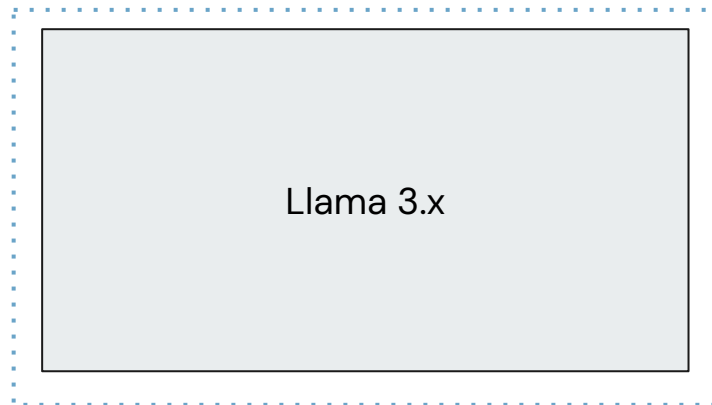
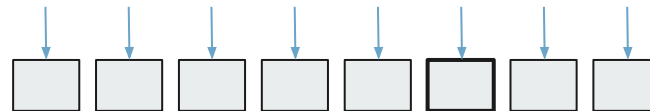
unisloth



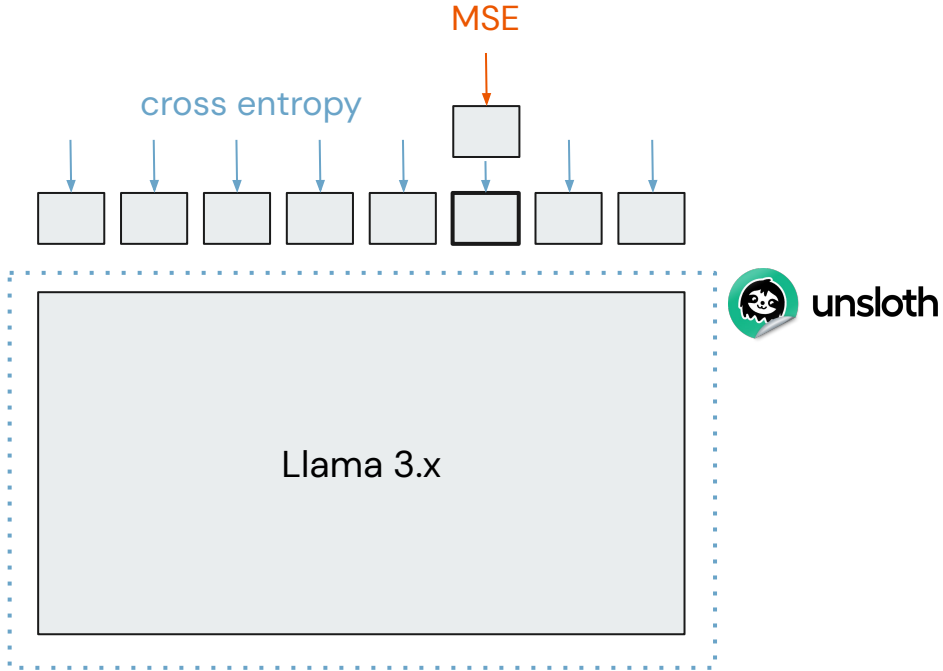


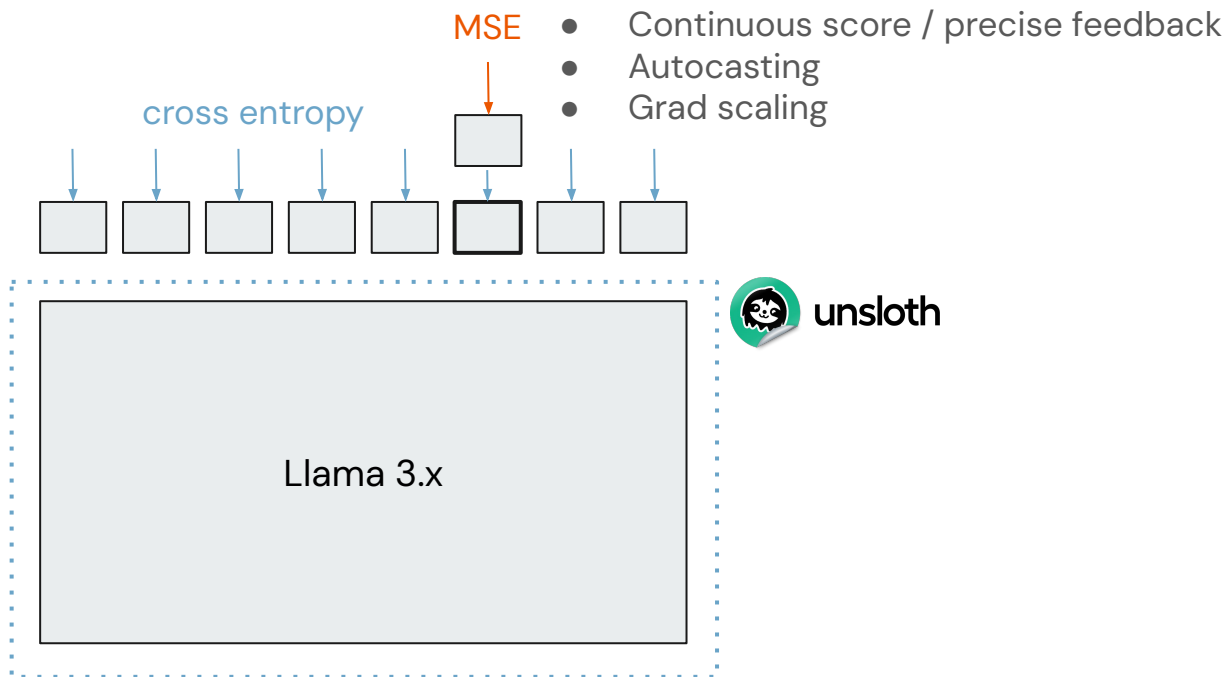


cross entropy



unisloth





Llama 3.x



Results



Configuration comparisons



Configuration comparisons

Llama Instruct

Llama 3.0

Instruction template

Zeroshot

Finetuned

Precise timing

Llama

Llama 3.1

Just EOT token

Few shot

NOT finetuned

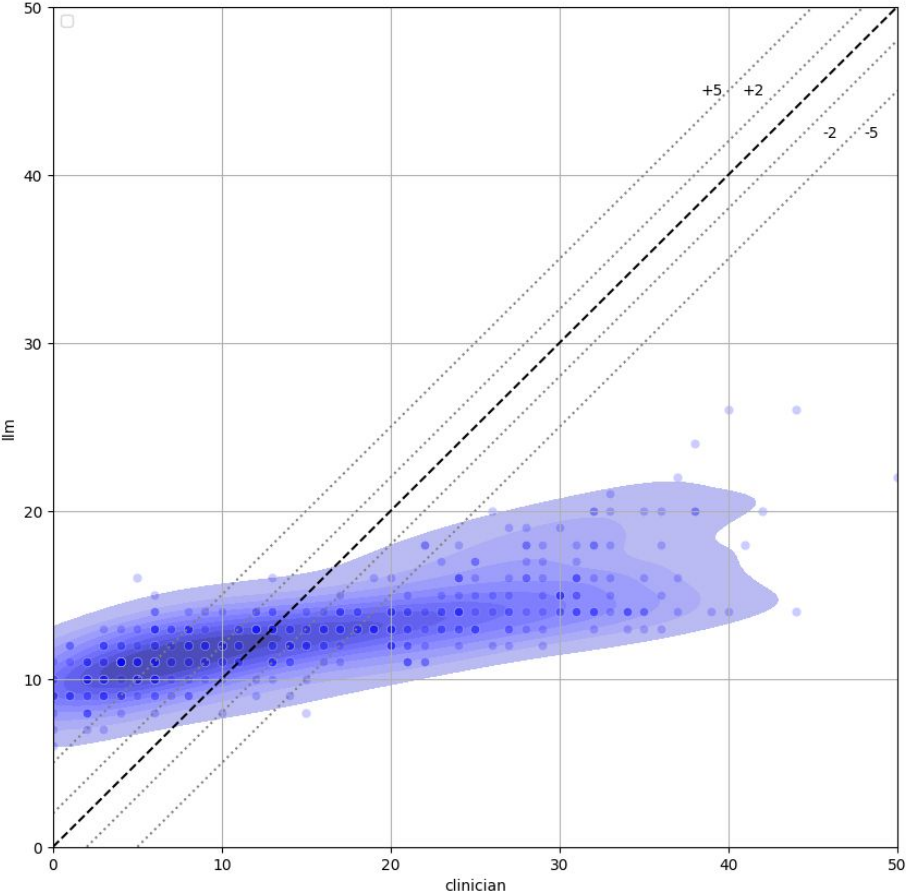
NOT precise timing



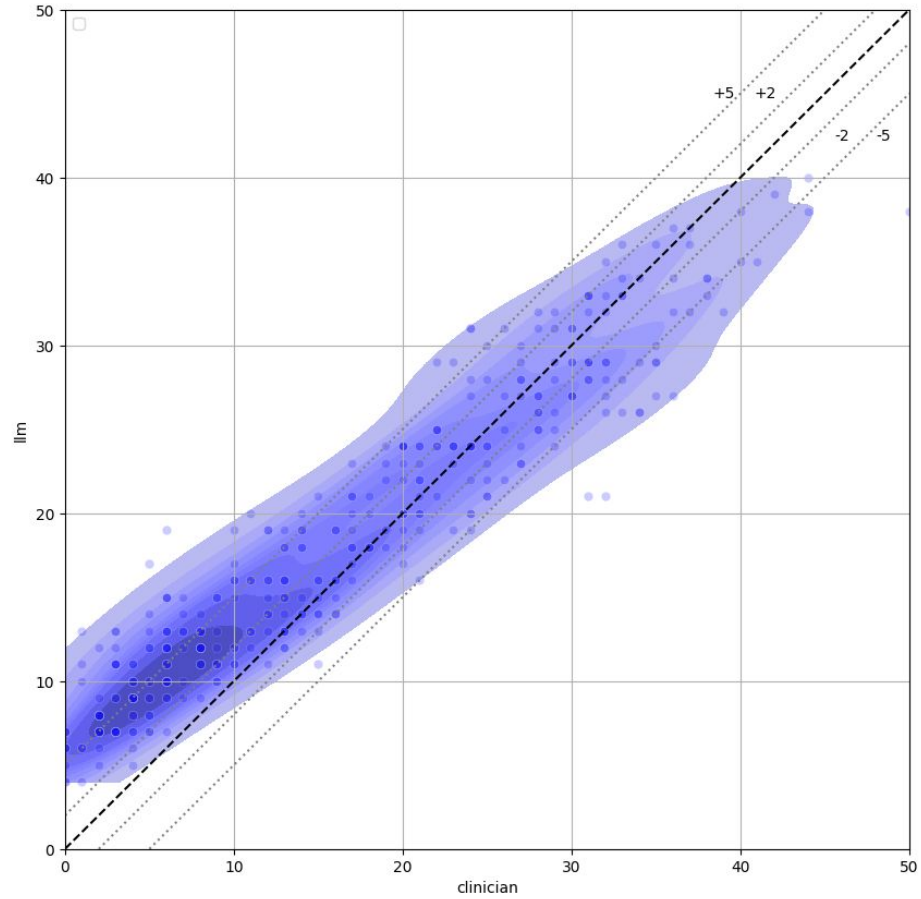
Configuration comparisons

Llama Instruct	>>	Llama
Llama 3.0	<	Llama 3.1
Instruction template	~	Just EOT token
Zeroshot	<	Few shot
Finetuned	>>	NOT finetuned
Precise timing	>>	NOT precise timing

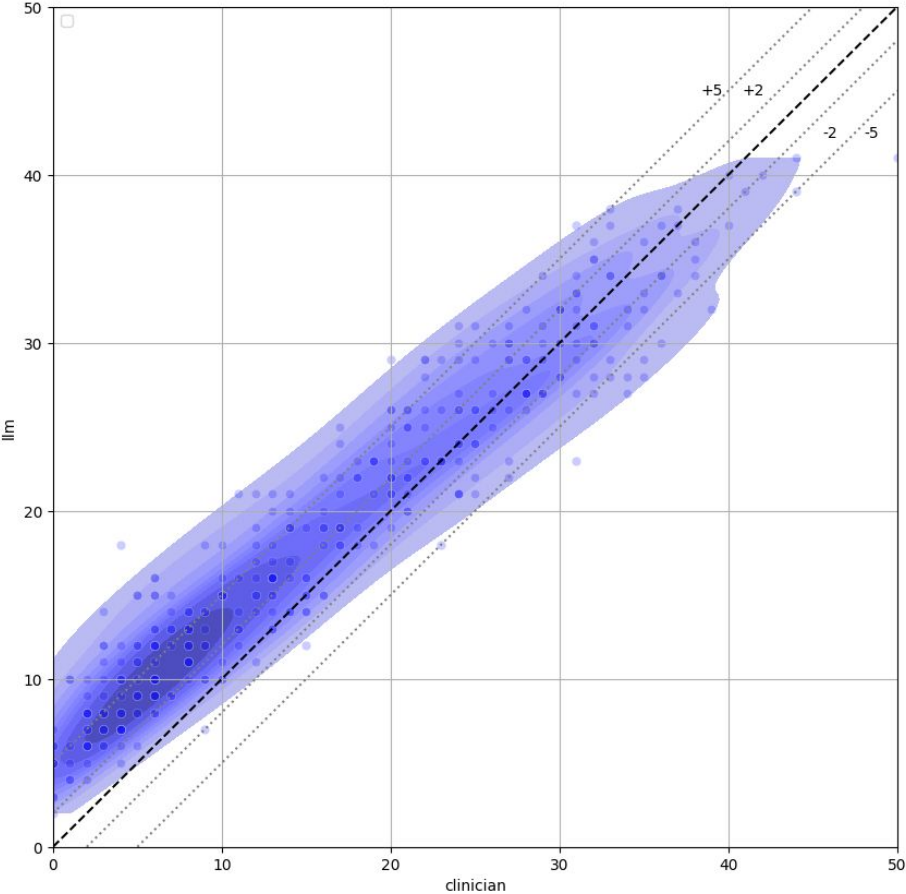
Zeroshot on Llama 3.1 8B
HAM-A score err: 8.13 +- 5.68, corr: 0.765



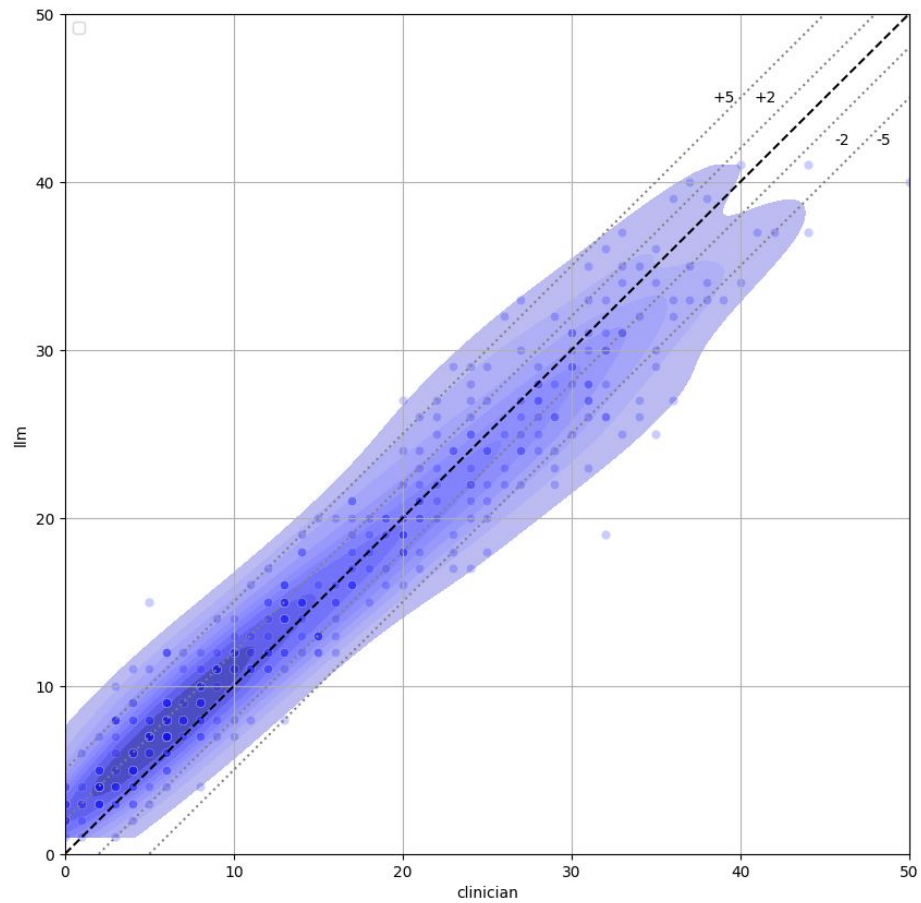
Zeroshot on Llama 3.1 8B Instruct
HAM-A score err: 3.77 +- 2.61, corr: 0.95



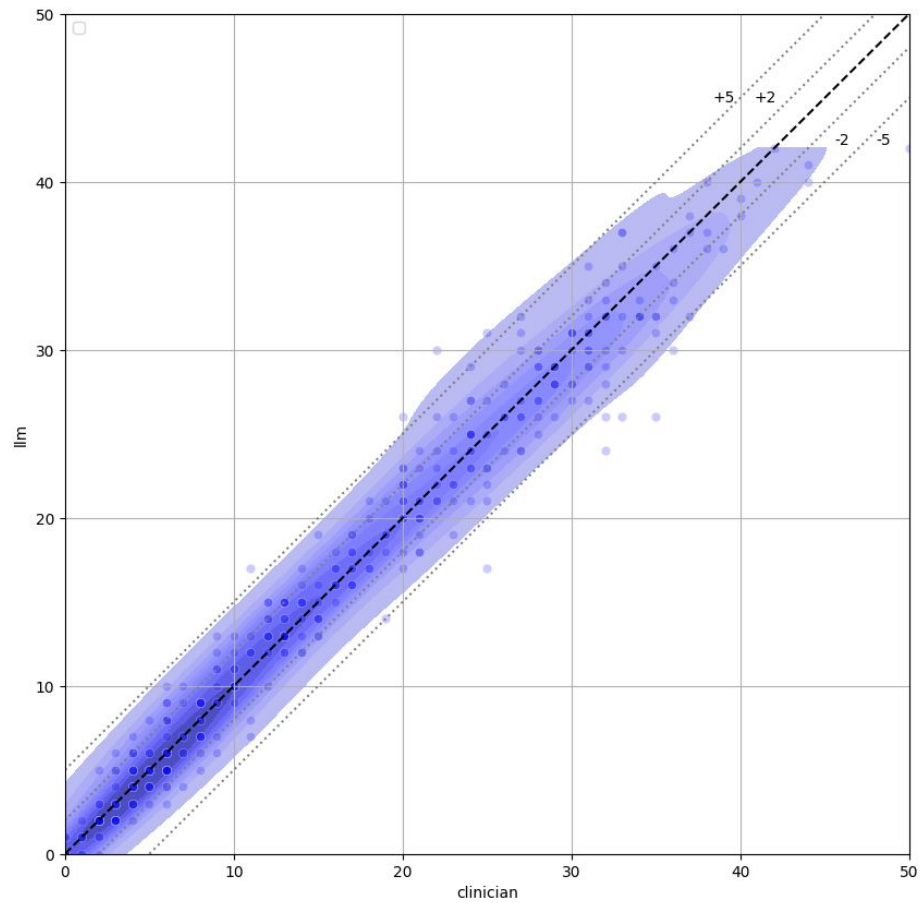
Fewshot on Llama 3.1 8B Instruct
HAM-A score err: 3.71 +- 2.46, corr: 0.957



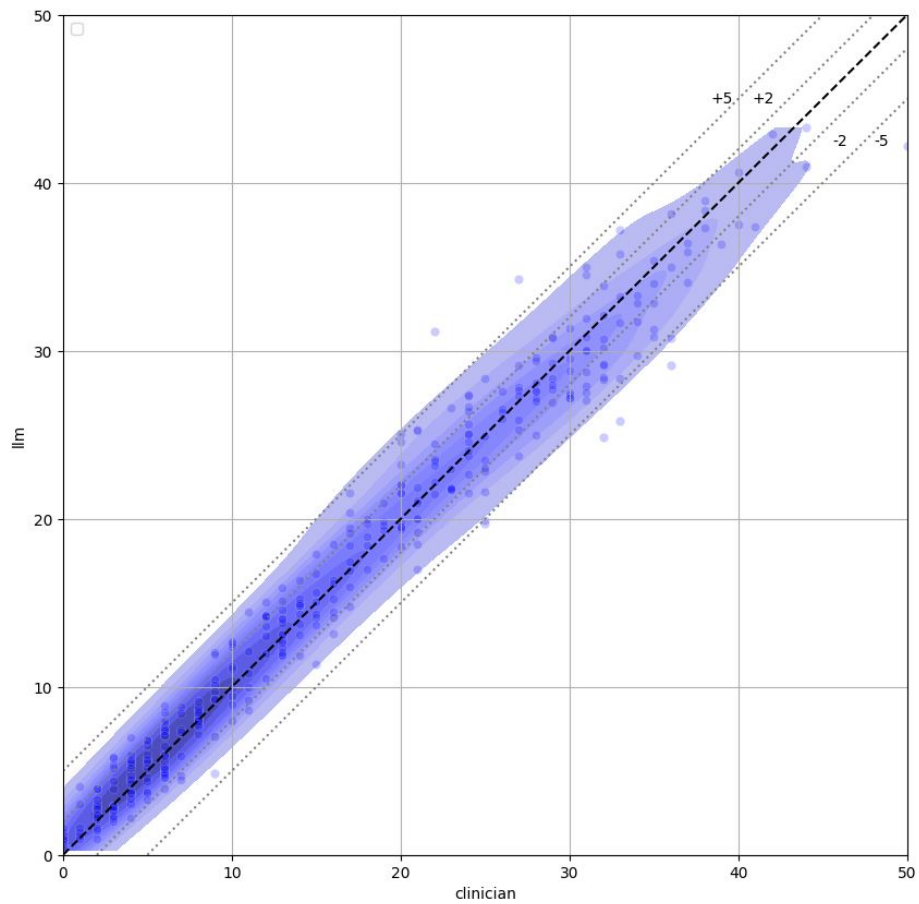
Fewshot (Long) on Llama 3.1 8B Instruct
HAM-A score err: 2.59 +- 2.03, corr: 0.961



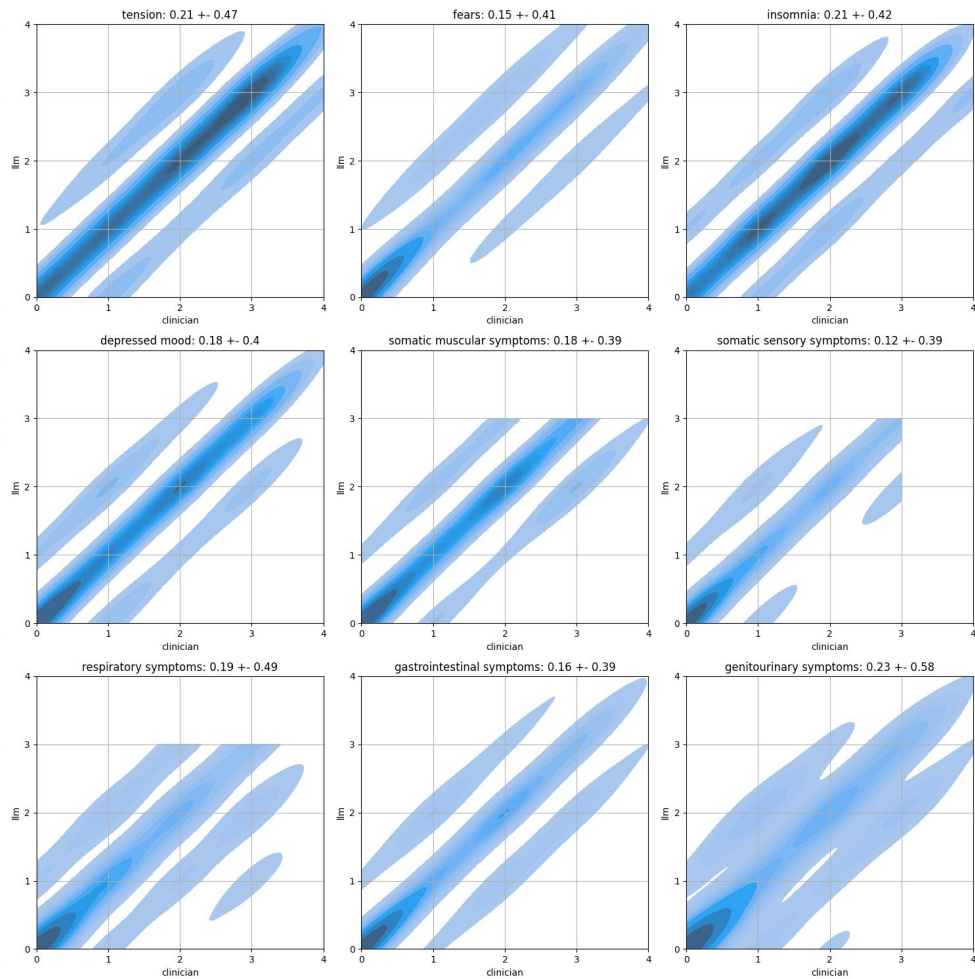
Finetuned Llama 3.1 8B Instruct
HAM-A score err: 1.56 +- 1.55, corr: 0.981



Finetuned Llama 3.1 8B Instruct with numeric output
HAM-A score err: 1.57 +- 1.39, corr: 0.983

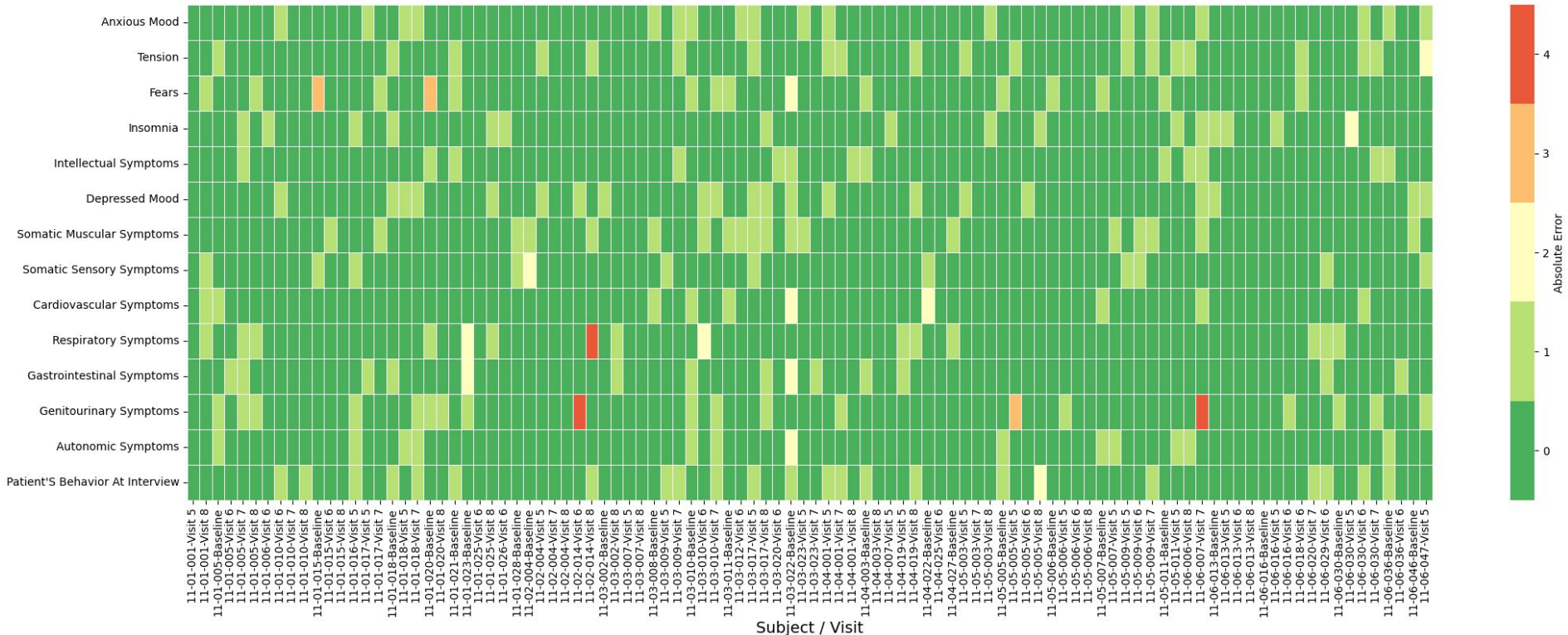


Clinical trial intelligence
#paradigma shift



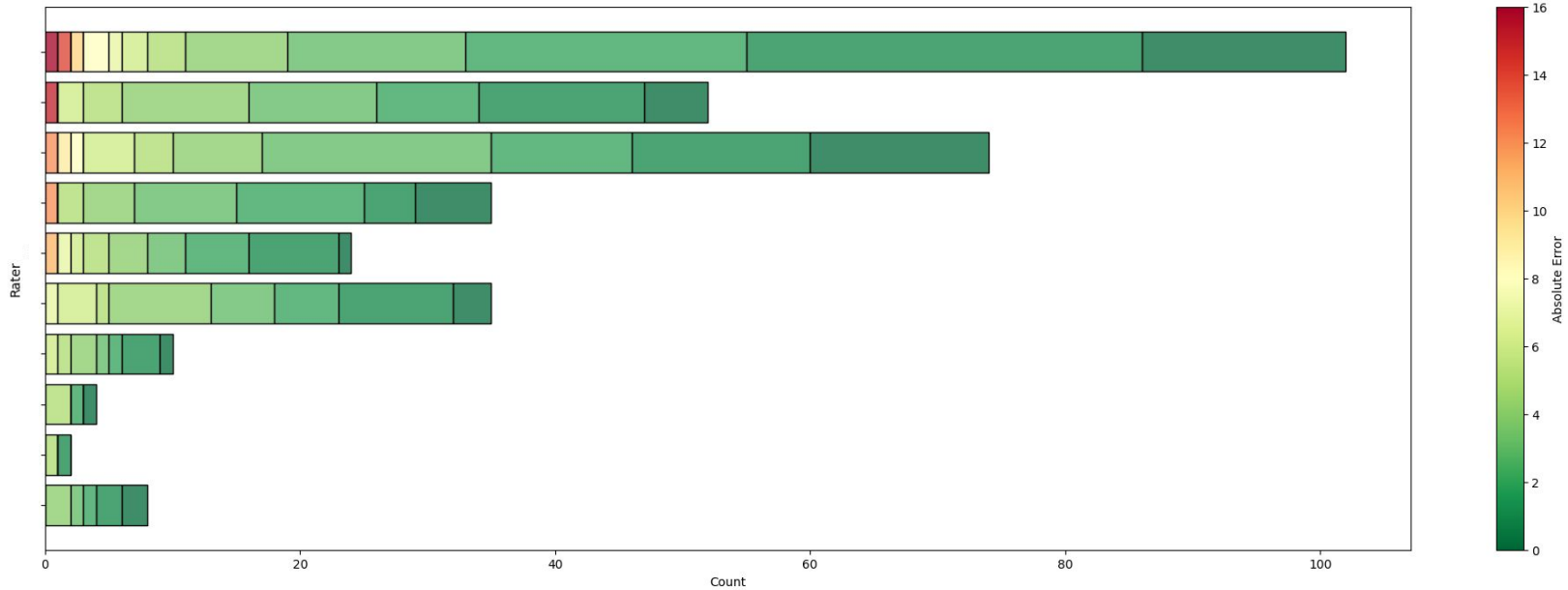
Big picture over individual symptoms may point to weak part of questionnaire that lead to interpretation with low inter rater agreement

HAM-A Item



Zooming into specific session and symptoms to find and interpret outliers.

Total accumulated errors



Tracking clinician performance during the time



Features

- Consistent ratings
- More nuanced continuous scales with confidence
- Concrete pointers to audio in case of FP/FN
- QA LLM for explainability
- Check whether clinicians follow the protocol
- Time spend on subitems
- Audio sentiment



Findings

- Inconsistent personal scales between clinicians
 - Difference between moderate and severe symptom
 - Over-rating 'very severe'
- Avoiding certain topics
 - sex / libido in genitourinary symptoms
- Different level of detail in questioning
 - Skipping questions
 - Changing the exact wording of questions
 - Chest pain / Chest pressure case
-



Findings

- Missing the facts (while making notes)
- Subject misunderstanding the question
 - asked about last week, describing last month
- Hard cases – unclear sources of symptoms
 - insomnia vs dog
 - headache vs chinese food/pills
 - lack of sex vs. menopause