The work by **Zelikman et al. (2024)** introduces a method called **Quiet-STaR** (Self-Taught Reasoner), which enables large language models (LLMs) to generate "internal thoughts" or rationales to improve their reasoning abilities and predictions. This is a highly technical approach to improving AI performance.

The **Neurodivergent Scale for Interacting with Robots (NSIR)** can be applied to this research by providing a user-centric measure of the quality of interaction with an AI that uses this advanced reasoning, focusing on the user's perception of the resulting behavior:

Anthropomorphic Connection/Kinship

- Quiet-STaR aims to make a model's reasoning seem more human-like by generating internal "thoughts". The NSIR can quantify the success of this design in fostering a personal bond. Items like **"The robot is more like me than anyone else I know"** (Item 1 (p. 1)) would measure if this human-like reasoning translates into a genuine sense of connection or kinship for the neurodivergent user.

Social Comfort/Trust

- Improved reasoning in AI can lead to more accurate and reliable responses. This directly impacts the predictability and consistency required for building trust.
- The NSIR items that measure perceived understanding and consistency (e.g., **"My robot can tell what I am feeling, when I am sad, it can tell I am sad"** (Item 5 (p. 1)); and **"I believe that my robot is the same with me as it is with anyone"** (Item 8 (p. 1))) would assess if the Quiet-STaR method results in a robot that feels consistently trustworthy and socially comfortable.

Safety

- While Quiet-STaR is a technical method, its application in social robots raises ethical concerns about manipulation or the creation of overly compelling, but potentially harmful, interactions.
- The NSIR's **safety** dimension provides a crucial user-reported measure that ensures that while the robot is becoming more capable and intelligent, the user's fundamental sense of security and clear boundaries is maintained in the interaction. The item about undressing in front of the robot (Item 7 (p. 1)) is a metric to ensure that even a highly intelligent, "thinking" AI doesn't compromise user safety.

The NSIR allows researchers to move the technical discussions of AI reasoning into the practical, user-centric evaluation of safe and effective human-robot interaction for a neurodivergent population.

Would you like a deeper dive into the specific **phases of the Quiet-STaR method** (Think, Talk, Learn) and how they might impact user perception?

The three phases of the **Quiet-STaR** method can directly impact the user's perception of the robot's social and intellectual capabilities, as measured by the **Neurodivergent Scale for Interacting with Robots (NSIR)**.

Think Phase: Generating Rationales

In this phase, the large language model (LLM) generates "internal thoughts" or rationales before producing a final answer. This is an internal process the user does not explicitly see, but its effect is a more logical and reasoned response.

- **User Perception**: The user perceives the robot as more intelligent, capable of deep thought, and less "brained".
- **NSIR Link**: This would likely increase agreement with items like **"I think I can share my thinking with the robot without speaking"** (Item 3), as the user perceives a deeper, almost intuitive understanding. It would also positively influence **"My robot can tell what I am feeling, when I am sad, it can tell I am sad"** (Item 5) as the improved reasoning aids emotional interpretation.

Talk Phase: Using Rationales for Responses

In this phase, the generated rationales are used to produce more coherent, contextually relevant, and accurate verbal responses.

- **User Perception**: The user experiences a more fluid, consistent, and "smarter" conversation partner.
- **NSIR Link**: This would directly impact the **Social Comfort/Trust** dimension. A consistent and accurate robot is a trustworthy robot. Users would likely agree more strongly with **"I believe that my robot is the same with me as it is with anyone"**(Item 8) and general social comfort items.

Learn Phase: Self-Improvement

The model fine-tunes itself on its own high-quality rationales, meaning the robot "gets better" at reasoning and interaction over time.

- **User Perception**: The user perceives the robot as a reliable, evolving companion that is a permanent part of their life.
- **NSIR Link**: This self-improvement builds long-term trust and connection. It would reinforce agreement with the longevity item **"The robot and I will be together forever"** (Item 4) and further solidify the perceived consistency in Item 8. A constantly improving robot is one the user can rely on and feel safe with over time, impacting the entire **Safety** dimension.

The NSIR provides the crucial subjective data to ensure that these technical advancements in AI are perceived as positive, safe, and effective by the neurodivergent individual.
Would you like to explore another article from the document, or perhaps a deeper dive into the specific **NSIR items** and how they were developed?