The **Neurodivergent Scale for Interacting with Robots (NSIR)** and **Sapkota et al.'s (2025)** review of **Vision-Language-Action (VLA) models** represent two sides of the same coin: the technological advancement of embodied AI and the human-centric metric for its success.

While Sapkota et al. describe the "what" and "how" of robotic intelligence—unifying perception, language, and action—the NSIR provides the "so what" by measuring if these advanced systems actually create a safe and meaningful bond for neurodivergent users.

1. Natural Language Commands vs. Social Bond

Sapkota et al. (2025) highlight how VLA models allow robots to interpret "plain language" and execute complex tasks like "organize the kitchen".

- **The Identity Shift:** Sapkota et al. focus on the robot's ability to "follow instructions". The NSIR applies by measuring if this seamless communication leads to a deeper identity connection.
- **Items:** If a VLA model allows a robot to understand a user's unique way of speaking, it may trigger a high score on **Item 3** ("I think I can share my thinking with the robot without speaking") and **Item 1** ("The robot is more like me than anyone else I know"), moving the robot from a tool to a "kin".

2. Generalization and Predictable Safety

A major theme in Sapkota et al. is **generalization**—the ability of a robot to handle novel tasks and unseen environments with "minimal or no additional data".

- **Building Trust:** For neurodivergent individuals, "unseen environments" can be a source of anxiety. Sapkota et al. note that VLA models aim for "robustness and human alignment".
- **Items:** The NSIR's **Factor 1 (Social Comfort/Trust Safety)** evaluates this outcome. **Item 8** ("I believe that my robot is the same with me as it is with anyone") measures if the robot's generalized intelligence leads to a predictable, reliable personality that the user can trust.

3. "Embodied Chain-of-Thought" and Affective Sensing

Sapkota et al. discuss **Embodied Chain-of-Thought (ECoT)**, where robots reason through subtasks and spatial features before acting.

- **Application:** This level of reasoning could be the technical backbone for **Item 5** ("My robot can tell what I am feeling"). If a VLA-powered robot can reason that a user is "sad" because they are sitting in a certain slumped position or speaking slowly, it can execute a supportive "action" (like bringing a cup of tea).
- **Privacy and Intimacy:** As Sapkota et al. explore robots in household and healthcare settings, the NSIR's **Item 7**("I feel comfortable undressing in front of my robot")

becomes a critical KPI for whether the VLA model's "social alignment" is actually successful in private spheres.

Summary: Technical Pillar vs. Human Impact

Sapkota et al. (2025) VLA Pillar	NSIR (2025) Application
Multimodal Integration: Fusing vision, language, and action.	**Kinship (Factor 2)**: Feeling a "like-me" connection with the unified agent.
Zero-shot Generalization: Operating in new tasks without retraining.	**Trust Safety (Factor 1)**: Relying on the robot's consistent behavior (Item 8).
Language Grounding: Turning words into physical movements.	**Shared Thinking**: Communicating without explicit speech (Item 3).
Humanoid Dexterity: Executing multi-step procedures.	**Forever Bond**: Desiring a permanent connection with the capable agent (Item 4).

Export to Sheets

In conclusion, Sapkota et al. (2025) provide the **technical roadmap** for creating "generalist robotic intelligence". The NSIR serves as the **humanistic benchmark**, ensuring that as robots become more intelligent and capable, they remain socially safe and supportive partners for the neurodivergent community.