UNIWERSYTET ZIELONOGÓRSKI

DYSKURS PRAWNICZY I ADMINISTRACYJNY 2/2026 ISSN 2657-926X

Martyna Łaszewska-Hellrigel
Uniwersytet Zielonogórski
ORCID: 0000-0002-2212-371X
m.laszewska-hellriegel@wpa.uz.zgora.pl

Between the Algorithm's Recommendation and the Individual's Will: Two Practical Case Studies on AI in Healthcare and Elder Care

Keywords: artificial intelligence, autonomy, informed consent, refusal of treatment, elder care, dignity of risk, public administration

Summary. This article reworks the discussion on artificial intelligence in healthcare and elder care from a predominantly conceptual debate into a case-oriented legal and bioethical analysis. Its central claim is that the most important practical risks do not arise where an AI system formally replaces the human decision-maker, but where it quietly restructures the conditions under which patients, professionals, relatives and care institutions act. The article therefore analyses two model case studies grounded in real and recurring patterns of practice: first, the refusal of life-sustaining treatment in the presence of an algorithmic recommendation for escalation of therapy; second, the gradual intensification of AI-supported home monitoring of an older person in the name of safety. In both cases, the decisive issue is not whether the technology is technically accurate, but whether legal and administrative arrangements preserve meaningful autonomy, contestability, proportionality and responsibility. The argument proceeds by connecting practical case analysis with medical-law principles concerning competent refusal, informed consent, dignity, privacy and the continuing relevance of the dignity of risk. The article concludes that a rights-sensitive use of AI requires procedural safeguards embedded in institutional design: functional transparency, documented dialogue, renewable consent, protected deviation from automated outputs, and accessible routes of contestation.

Paternalizm algorytmiczny czy upodmiotowiona autonomia? Teoretycznoprawne refleksje na temat sztucznej inteligencji w opiece nad osobami starszymi i problematyce zgody

Słowa kluczowe: sztuczna inteligencja, autonomia, świadoma zgoda, odmowa leczenia, opieka nad osobami starszymi, godność ryzyka, administracja publiczna

Streszczenie. W miarę jak systemy sztucznej inteligencji zyskują na znaczeniu w ochronie zdrowia i usługach społecznych, ich zastosowanie w opiece nad osobami starszymi skłania do ponownego podjęcia teoretycznoprawnej debaty na temat autonomii, świadomej zgody oraz ewoluującej roli prawa w regulowaniu wspomaganego maszynowo podejmowania decyzji. Artykuł analizuje normatywne implikacje narzędzi algorytmicznych stosowanych w opiece nad osobami w podeszłym wieku, ze szczególnym uwzględnieniem napięcia między ochroną autonomii jednostki a promowaniem jej dobrostanu za pośrednictwem systemów zautomatyzowanych.

Koncentrując się na środkowoeuropejskich porządkach prawnych – w tym Polski i Słowacji – w szerszym kontekście prawa unijnego, artykuł analizuje, w jaki sposób oparte na AI systemy monitorowania, triaż i rekomendacji mogą wpływać na zdolność prawną, zastępcze podejmowanie decyzji oraz wykonywanie woli przez osoby starsze. Szczególną uwagę poświęcono modelom zgody w zakresie dawstwa narządów (*opt-in* i *opt-out*) oraz dyrektywom opiekuńczym, badając, czy algorytmicznie kształtowane rozwiązania domyślne osłabiają, czy też wspierają rzeczywisty wybór.

Artykuł odwołuje się do prawa porównawczego, bioetyki oraz teorii autonomii relacyjnej, aby ocenić, kiedy AI stanowi formę paternalizmu, a kiedy może służyć jako narzędzie prawnego upodmiotowienia. Kończy się rekomendacjami dotyczącymi włączenia zasad przejrzystości, proporcjonalności i kontroli użytkownika do standardów prawnych regulujących algorytmiczne podejmowanie decyzji w opiece nad osobami starszymi, postulując wyważoną równowagę między efektywnością technologiczną a poszanowaniem godności człowieka.

1. Introduction

Artificial intelligence now appears in healthcare and care administration not as a distant prospect, but as part of the ordinary infrastructure of decision-making. Risk prediction tools, triage systems, diagnostic support interfaces, remote monitoring solutions, care-navigation applications and behaviour-detection systems are increasingly presented as instruments of assistance. Their promise is familiar: they are said to improve accuracy, reduce human error, allocate scarce resources more efficiently and make care more continuous, anticipatory and personalised. Yet, the legal and bioethical challenge begins precisely at the point where the language of assistance sounds most reassuring. An instrument can assist without being neutral. It can support a decision while also shaping the field in which that decision is made. In contemporary data-rich environments, that shaping power becomes especially strong because the system does not merely display information; it sorts, highlights, ranks, alerts, prompts and normalises[1].

For that reason, the central question is not simply whether AI 'decides'. In many practical settings, it does not decide in a formal sense. A physician still signs the order, a patient still gives or refuses consent, a family member still approves the installation of a monitoring device, and a care institution still adopts the policy. But between formal decision-making power and real normative influence, there is a wide and legally important space. The architecture of options can be curated in advance; what counts as prudent, safe, explainable or professionally defensible can be pre-structured; and the burdens of justification can be redistributed so that deviating from the machine's recommendation becomes institutionally costly. This is why current debates on autonomy cannot stop at the classical image of a face-

[1] K. Yeung, *'Hypernudg": Big Data as a mode of regulation by design, Information*, "Communication & Society" 2017, vol. 20(1), pp. 118-136.

to-face encounter between two human agents. Autonomy must also be analysed at the level of design, workflow and governance[2].

This article proceeds from the view that case-based analysis is particularly useful for this task. General discussions about 'autonomy', 'fairness' or 'human oversight' are important, but they often remain too abstract to show how pressure is actually generated in practice. The crucial transformations occur in concrete situations: when a competent patient refuses a beneficial intervention; when the record is written in a way that casts refusal as irrational; when a nurse feels unable to contradict a risk score; when an older person who initially agreed to a simple alert system gradually becomes the object of continuous surveillance; or when relatives invoke safety to legitimise a level of control that alters the meaning of home itself. In those moments, law and bioethics must move from slogans to criteria[3].

The article therefore centres on two practical case studies. They are modelled cases rather than reports of one single lawsuit, but each is firmly rooted in patterns that are already visible in medical practice, elder care, scholarship on automation bias and the growing literature on digital surveillance in care settings. The first case concerns a competent patient's refusal of invasive treatment in circumstances where an AI-supported clinical system strongly recommends escalation. The second concerns an older person living at home whose initially limited monitoring arrangement is repeatedly expanded in the name of safety. Both cases are analytically productive because they involve a tension between will and welfare, but they do so in different institutional forms. In the first, the conflict appears acutely and dramatically at the bedside. In the second, it unfolds gradually through a series of apparently small administrative and technical adjustments.

The argument developed below is straightforward but demanding in its implications. AI does not threaten autonomy, only when it openly replaces human judgement. It also threatens autonomy when it silently alters the practical conditions of choice, when it privileges one understanding of rationality over others, and when institutions organise accountability in ways that reward compliance with the system and penalise justified resistance. From that perspective, the most important legal task is not merely to ask whether AI may be used, but under what procedural conditions its use remains compatible with the rights, dignity and agency of the person whose life is being governed by the system.

---

[2] R.H. Thaler, C.R. Sunstein, *Libertarian Paternalism*, "American Economic Review" 2003, vol. 93(2), pp. 175-179; K. Yeung, *'Hypernudge': Big Data as a mode of regulation*…, pp. 118-136.

[3] O'Neill's critique remains relevant here because formal consent may coexist with structurally weak autonomy: O.O'Neill, *Autonomy and Trust in Bioethics*, Cambridge 2002, p. 49.

2. Why a Case-oriented Approach is Needed

A case-oriented approach is valuable because AI governance in healthcare and elder care is rarely exhausted by questions of technical performance. A prediction can be statistically impressive and still produce a legally distorted encounter. A monitoring system can detect risk accurately and still damage privacy, agency and trust. Conversely, a system may be imperfect yet remain acceptable if its institutional use preserves contestability, limits intrusiveness and leaves room for humanly intelligible reasoning. The law therefore needs categories that are sensitive not only to outcomes, but also to the way in which outcomes are reached.

The first conceptual point concerns autonomy. In medical law, respect for autonomy has never meant that a competent person must choose what others consider objectively best. It means, rather, that a person who can understand relevant information, deliberate and communicate a decision retains authority over certain deeply personal choices, including choices that others regard as mistaken or dangerous. The problem introduced by AI is that systems built on statistical optimisation can subtly recode this relationship. Once a system defines the medically recommended path with unusual confidence, refusal may cease to appear as one legitimate option, among others, and instead come to be treated as a deviation from the rational baseline[4].

The second point concerns rights theory. A will-based understanding of rights highlights the importance of normative control: the point of a right is that the right-holder can decide whether to authorise, refuse, waive or insist. An interest-based understanding emphasises the protection of welfare, well-being and objective interests. Both perspectives illuminate AI-supported care. If one looks only through the lens of welfare, algorithmic steering may seem justified whenever it improves measurable outcomes. If one looks only through the lens of will, one may fail to appreciate genuine vulnerability, dependency and cognitive burden. The practical challenge is therefore not to choose one theory once and for all, but to prevent welfare-oriented design from quietly erasing the person's normative authority[5].

The third point concerns rationality. Healthcare law traditionally protects competent decisions even when they are not medically prudent, because competence is not identical with substantive wisdom. Yet, AI systems are built precisely to predict, compare and optimise. They derive authority from the capacity to identify the option most strongly associated with survival, adherence, reduced cost or re-

[4] P.S. Appelbaum, *Assessment of Patients' Competence to Consent to Treatment*, "New England Journal of Medicine" 2007, vol. 357(18), pp. 1834-1840.

[5] J. Raz, *Legal Rights*, "Oxford Journal of Legal Studies" 1984, vol. 4(1), pp. 1-21.

duced risk. That makes them powerful tools, but also potentially paternalistic ones. Once optimisation criteria are embedded in workflow, rationality can migrate from a procedural notion-did the person understand and decide? To a substantive one-did the person choose what the model identified as the most beneficial option? [6]

Case studies expose this migration in a way that abstract principle cannot. They show where the recommendation appears, who sees it, how it is documented, how dissent is framed, who bears the burden of explanation and what institutional incentives are at work. They thus make visible an often neglected truth: many rights violations in technologically mediated care are not produced by a dramatic single act, but by the cumulative interaction of interface design, professional psychology, family anxiety and organisational routines.

3. Case Study One: Refusal of Treatment Against an Algorithmic Recommendation

3.1. Factual Structure of the Case

A patient with severe respiratory failure is admitted to hospital. He is alert, oriented and able to communicate consistently. He is informed that his condition may deteriorate and that invasive ventilation would likely prolong life. The hospital uses an AI-supported clinical deterioration system trained on prior patient data. On the basis of physiological variables and comparable trajectories, the system assigns the patient a very high-risk score and generates a prominent recommendation that escalation of treatment should not be delayed. The treating team, already under pressure because intensive-care beds are limited and delayed decisions are closely reviewed, receives repeated alerts. The patient nevertheless states, calmly and repeatedly, that he does not want intubation. He accepts oxygen, symptom relief and further conversation, but refuses invasive ventilation.

At first glance, the legal structure of the case appears familiar. It resembles the classical problem of a competent refusal of life-sustaining treatment. Yet, the presence of the algorithm changes the normative atmosphere. The physician is no longer merely dealing with a clinical disagreement or a tragic conflict between prognosis and patient will. The refusal now stands against a quantified, repeatable and institutionally visible recommendation produced by a system marketed as reducing error and delay. The machine does not have legal personality and does

[6] On the risk that digitally mediated environments turn influence into design-based steering, see K. Yeung, *'Hypernudge': Big Data as a mode of regulation by design, Information,* "Communication & Society" 2017, vol. 20(1), pp. 118-136.

not consent on anyone's behalf. Nevertheless, it exerts pressure by changing what counts as responsible action inside the organisation.

The most significant feature of the case is therefore not the alert itself but the redistribution of justificatory burden. Before the introduction of the system, the patient's refusal had to be taken seriously because the clinician needed positive reasons to override it, and those reasons would normally be unavailable where capacity is intact. After the introduction of the system, however, the clinician who accepts the refusal may feel required to justify not only her own judgement but also her departure from the algorithmic output. The patient's will, while formally unchanged, becomes practically harder to respect.

3.2. Legal Analysis

Medical law has long treated competent refusal as a core expression of bodily autonomy. The underlying principle is not that refusal is always wise, but that invasive treatment without valid consent constitutes a profound interference with bodily integrity. Capacity is thus the threshold question, not agreement with medical advice. The standard view in contemporary medical law is that an adult patient with decision-making capacity may refuse treatment even where the likely consequence is death. That proposition would lose much of its meaning if technological systems could indirectly downgrade the refusal by branding it as statistically irrational or institutionally non-compliant[7].

The case law on refusal and consent confirms the same orientation. In Re B, the English court recognised that a competent adult may refuse life-sustaining treatment even when others consider the decision tragic or misguided. Later jurisprudence on informed consent, including Montgomery, reinforced the broader move away from professional paternalism by insisting that the patient is not merely the object of medical expertise but a legal agent entitled to information material to her own decision. These authorities are not about AI, but they become highly relevant once algorithmic recommendations begin to alter the informational and institutional environment in which consent and refusal are recorded[8].

From this perspective, the legal problem is not whether the model may be consulted. It usually may. The problem is whether consultation quietly becomes normative displacement. Several pathways of displacement deserve particular attention. First, documentation may subtly change its vocabulary: a refusal may be

[7] P.S. Appelbaum, *Assessment of Patients' Competence to Consent to Treatment*, "New England Journal of Medicine" 2007, vol. 357(18), pp. 1834-1840.

[8] Montgomery v Lanarkshire Health Board [2015] UKSC 11; Re B (Adult: Refusal of Medical Treatment) [2002] 2 All ER 449; see also M. Stauch, *Comment on Re B (Adult: Refusal of Medical Treatment) [2002] 2 All England Reports 449*, "Journal of Medical Ethics" 2002, vol. 28(4), pp. 232-233.

described less as an autonomous choice and more as non-compliance with the recommended pathway. Secondly, professional discussions may begin from the assumption that the system's output expresses the most responsible option, so that respect for refusal appears as an exception. Thirdly, retrospective review mechanisms may reward adherence to the alert and expose deviation to internal scrutiny. Under those conditions, the right formally survives while its exercise becomes structurally disfavoured.

A further issue is the quality of information disclosed to the patient. If the clinician tells the patient simply that 'the computer says your risk is very high', the patient may be exposed to a particularly authoritative form of persuasion without receiving an intelligible account of what the system actually does, what kind of data it relies on, how confident the output is, or whether comparable clinicians sometimes disagree with it. The informational asymmetry becomes deeper because the aura of objectivity attached to the model may be greater than the patient's ability to interrogate it. Respect for autonomy in such conditions requires more than passing on the alert. It requires translating the role of the system into terms the patient can meaningfully use.

Legally, then, the crucial point is that capacity, consent and refusal remain person-centred categories. They do not collapse into algorithmic optimisation. A hospital that uses AI in a manner that effectively penalises clinicians for honouring a competent refusal risks reintroducing paternalism in technologically updated form. It also risks blurring responsibility: if the intervention proceeds against the person's will, the relevant legal wrong cannot be neutralised by the fact that the system recommended it.

3.3. Bioethical Analysis

The bioethical difficulty of the case lies in the tension between welfare and authority. The system may indeed predict that ventilation would prolong life; the clinicians may sincerely believe that treatment is in the patient's best interests; relatives may be distressed by the prospect of a refusal. None of that is ethically irrelevant. But beneficence does not erase agency. Respect for persons includes accepting that a competent individual may evaluate burdens, dependence, fear, invasiveness, future quality of life and personal meaning differently from the clinical team. To insist that the medically beneficial option must prevail is to replace moral dialogue with a narrower idea of optimisation[9].

[9] For a concise account of the tension between autonomy, individuality and consent, see O.O'Neill, *Autonomy and Trust in Bioethic*s, Cambridge 2002, p. 40.

This is where the notion of trust becomes decisive. Trust in healthcare is not strengthened when the professional appears merely to transmit an opaque recommendation backed by institutional pressure. It is strengthened when the patient can see that technology is being used as an aid to conversation rather than as a device that settles the matter in advance. O'Neill's critique of thin and over-individualised conceptions of autonomy remains useful here: autonomy cannot be secured by ritualised consent talk if the surrounding structure renders disagreement practically futile[10].

The case also illuminates a familiar cognitive danger. Human operators often over-rely on automated systems, especially when those systems are presented as safety-enhancing and when institutional cultures treat them as markers of diligence. Research on human-automation interaction has repeatedly shown that reliance on automation can generate both omission and commission errors: people may fail to act when the system does not prompt them, and they may follow the prompt, even when independent judgement should lead elsewhere. In the clinical setting, that dynamic does not merely threaten technical accuracy. It can also transform the ethical stance of the clinician by making resistance appear reckless[11].

Accordingly, the ethical failure in this case would not consist only in physically imposing treatment. It could already occur earlier, at the point where the conversation becomes subtly coercive, where refusal is repeatedly reclassified as confusion, where documentation narrows the patient's reasons into a problem of non-adherence, or where clinicians are made to feel that honouring refusal is professionally indefensible. AI magnifies these risks because it gives optimisation a visible interface and an administrative trace.

3.4. Administrative Implications

For public administration and institutional governance, the first case demonstrates that human oversight cannot be reduced to the mere presence of a human final signatory. Oversight has to be organised. A clinician must be able to depart from the model's recommendation without entering a zone of institutional suspicion. Documentation systems should therefore record not only that an alert was generated, but also that competent refusal remains legally decisive and that deviation from the algorithmic pathway may be justified by patient choice. Hospitals that

[10] O.O'Neill, *Autonomy and…*, pp. 44.

[11] R. Parasuraman, V. Riley, *Humans and Automation: Use, Misuse, Disuse, Abuse*, "Human Factors" 1997, vol. 39(2), pp. 230-253; K. Goddard, A. Roudsari, J.C. Wyatt, *Automation bias - a hidden issue for clinical decision support system use*, "Studies in Health Technology and Informatics" 2011, vol. 164, pp. 17-22; L.J. Skitka, K.L. Mosier, M. Burdick, *Does automation bias decision-making?*, "International Journal of Human-Computer Studies" 1999, vol. 51(5), pp. 991-1006.

fail to build such space into workflow risk producing a de facto hierarchy in which the model outranks the patient.

The case also shows why the language of explanation matters. Institutions should not confine themselves to broad statements that AI is 'used to support decisions'. In disputes about consent and refusal, the patient and the record should reflect the actual function of the system: whether it predicts deterioration, whether it recommends a threshold for escalation, whether it influences triage order, and whether clinicians are expected to document disagreement. Functional transparency is a precondition of meaningful contestation.

4. Case Study Two: AI-supported Home Monitoring and the Gradual Expansion of Control

4.1. Factual Structure of the Case

An older woman lives alone in her own home. She has early cognitive impairment but remains able to manage most daily activities and clearly states that remaining at home matters deeply to her. Her daughter, worried after two minor incidents of disorientation, arranges a home-monitoring package through a care provider. At the outset, the system is modest: a door sensor, medication reminders and an alert if no movement is detected for an unusual period. The older woman agrees, partly because she wants to reassure her family and avoid institutionalisation.

Over the next months, the arrangement changes. The provider introduces predictive analytics that classify patterns of activity as normal or abnormal. Nighttime wandering alerts are activated. Audio prompts are added. Then the daughter requests camera-based monitoring in the hallway, arguing that she needs to know whether her mother has fallen. The provider explains that the upgraded package is safer and easier to manage. The older woman expresses discomfort, saying that she feels watched, but is told that the changes are necessary because her risk profile has worsened. Nothing dramatic happens in a single step. Rather, a series of individually defensible adjustments gradually transforms the home into a monitored environment governed by risk detection.

This second case differs from the first in tempo and tone. There is no single bedside refusal and no immediate life-or-death confrontation. Yet, the normative stakes are no less serious. What is at issue is whether safety can become a solvent that dissolves privacy, spontaneity and the practical authority of the older person over the terms of her own everyday life. The case is especially important because the justifications for intervention are plausible, emotionally compelling and often motivated by genuine concern.

4.2. Legal Analysis

The legal significance of the case lies first in the continuing relevance of consent, and secondly in the limits of treating an initial consent as a blanket authorisation for future escalation. Agreement to a basic assistive arrangement does not automatically legitimise every later increase in intrusiveness. When the technology changes function-moving from reminders to behavioural inference, from passive sensing to camera-based observation, or from emergency response to continuous pattern analysis-the normative object of consent also changes. The person's earlier agreement cannot simply be stretched to cover the new system.

Data protection law helps to clarify this point, even though the issue is broader than data governance alone. Principles such as purpose limitation, data minimisation and proportionality point against the easy expansion of surveillance whenever more data become technically available. In a home-care context, moreover, personal data are inseparable from lived experience. The hallway camera does not merely 'process data'; it modifies the conditions of domestic life, alters the sense of being alone, and changes the relation between dependence and intimacy. Legal analysis must therefore avoid the reductionist error of treating surveillance only as an information-management problem[12].

The case also raises an important issue of capacity and vulnerability. Early cognitive impairment may justify supportive communication, repeated discussion and tailored safeguards. It does not justify the assumption that the person's present preferences have lost normative significance. Too often, systems of care slide from 'supporting decision-making' to 'pre-empting objection'. The mere invocation of safety is insufficient to justify that slide. If the person understands the relevant change well enough to express resistance to more intrusive monitoring, that resistance must remain a legally meaningful fact.

Finally, the institutional role of family members and providers requires scrutiny. Relatives frequently act out of fear, guilt and responsibility. Care providers often operate in environments where adverse events trigger blame and where surveillance appears as a defensible risk-management tool. The law should recognise those pressures without letting them displace the subject of care herself. Otherwise, home care becomes an administrative arrangement negotiated around the older person rather than with her.

[12] European Parliament Regulation (EU) 2016/679 Council of 27 April 2016 (General Data Protection Regulation).

4.3. Bioethical Analysis

The bioethical core of the second case is captured by the idea of the dignity of risk. The phrase is old, but its relevance has grown rather than diminished. Perske's classic insight was that overprotection can degrade persons by depriving them of the chance to make choices that carry some measure of risk. Contemporary debates in elder care have restated the same thought in rights-based terms: if all significant risk is designed out of life, what remains may be physically safer but existentially diminished[13][14].

AI surveillance technologies intensify this dilemma because they permit an unprecedented continuity of observation. The question is no longer only whether a care worker checks on someone too often. It is whether the background environment itself becomes interpretive and interventionist-constantly sorting behaviour into risk categories, generating alerts, prompting relatives to intervene and nudging the person toward pre-approved patterns of conduct. Such systems can be defended as benign because they often operate quietly. That quietness is precisely what makes them ethically powerful.

The language of safety should therefore be handled with care. Safety is not a value that automatically trumps all others. In the life of an older person who wishes to remain at home, values such as privacy, routine, intimacy, solitude, unpredictability, self-presentation and the ordinary right not to be constantly assessed may be central to well-being. A morally serious analysis cannot define welfare in purely actuarial terms. Joseph's recent argument that elder-care technologies should begin from a commitment to dignity rather than mere efficiency is especially apt in this respect[15].

The case also reveals the ethical inadequacy of one-off consent. In dynamic monitoring environments, the person does not confront a single intervention but an evolving regime. Ethical legitimacy must therefore be renewable. Every increase in intrusiveness changes the balance between protection and self-determination. To insist on renewed discussion is not bureaucratic formalism; it is the practical expression of respect for agency under conditions of technological drift.

---

[13] M. Foundas, *Dignity of risk in residential aged care: a call to reframe understandings of risk*, "Medical Journal of Australia" 2025, vol. 223(4), pp. 187-188.

[14] R. Perske, *The Dignity of Risk and the Mentally Retarded*, "Mental Retardation" 1972, vol. 10(1), pp. 24-27.

[15] J. Joseph, *Designing for dignity: ethics of AI surveillance in older adult care, Frontiers in Digital Health* 2025, vol. 7, article 1643238.

4.4. Administrative Implications

Administratively, the second case shows that care providers need escalation rules tied to proportionality rather than to mere technical capability. The fact that the system can infer more, store more or show more does not answer whether it should. Providers should therefore separate emergency functions from lifestyle surveillance, distinguish fall detection from behaviour scoring, and document why a less intrusive option is insufficient before migrating to a more invasive one. Without such discipline, AI-supported home care becomes vulnerable to what may be called incremental normalisation: each addition looks minor, but together they produce a different moral world.

Equally important is the availability of contestation. Older persons and their relatives should know who has authority to change the settings, how those changes are reviewed, and how disagreement is handled. This is where public-administration concerns and private-care practices meet. Even where care is delivered through mixed or private arrangements, the governance logic of rights, proportionality and review remains indispensable. Otherwise, the most consequential decisions are hidden inside procurement choices, service packages and dashboard settings rather than treated as matters of rights-bearing significance.

5. Cross-case Lessons for Law and Public Administration

The two cases are institutionally distinct, yet they converge on a single lesson: the central legal question is not whether AI is useful, but whether the surrounding procedure preserves meaningful human agency. In the first case, the threat arises from the authority of optimisation in a high-stakes clinical encounter. In the second, it arises from the slow sedimentation of surveillance into everyday care. But in both, the decisive danger is the same. The person remains formally present as decision-maker while the substantive terms of the decision are increasingly organised elsewhere.

Several safeguards follow from this insight. The first is functional transparency. Individuals should be told, in intelligible terms, what role the system plays in the relevant decision. Not every person needs a technical explanation of model architecture; what matters is whether the person can understand the practical function of the system in the pathway that affects her. In the treatment-refusal case, that means explaining that the system predicts deterioration and prompts escalation. In the home-monitoring case, it means explaining whether the technology merely alerts after a fall or also classifies patterns of behaviour. Functional opacity is incompatible with meaningful self-determination.

The second safeguard is documented dialogue rather than documentation of compliance. Records should show that the person's reasons were heard, not simply that the recommended pathway was communicated. In clinical settings, this means recording the patient's values, burdens and expressed priorities, not only the risk score. In care settings, it means documenting objections to monitoring changes and the reasons why less intrusive alternatives were considered sufficient or insufficient. Documentation that only mirrors the system's categories distorts the normative landscape.

The third safeguard is renewable consent tied to changes in intrusiveness. This is particularly important in elder care, but it also matters in healthcare whenever tools are repurposed or extended. A person may agree to one form of assistance and reasonably object to another. Institutions should therefore treat material functional expansion as a new normative event, rather than as a mere update of the same service. This point aligns both with rights-sensitive care ethics and with legal principles that resist overbroad consent[16].

The fourth safeguard is protected deviation from automated outputs. Human oversight is real only if a professional can disagree without triggering a presumption of negligence. Where every deviation from the alert requires exceptional justification, the system effectively governs. Healthcare institutions and care providers should instead normalise reasoned departure. The requirement should be explanation, not obedience. This approach is also more consistent with current European regulation, which increasingly treats human oversight, accountability and risk management as design obligations rather than rhetorical aspirations[17].

The fifth safeguard is accessible contestation. Persons affected by AI-supported decisions need a route by which they can object, seek review and ask for reconfiguration. In public administration, procedural justice depends not only on the initial decision but also on the availability of challenge. In technologically mediated care, this means that rights should not evaporate into software settings. If a patient objects to the algorithmic framing of refusal, or an older person objects to camera-based monitoring, there must be a humanly accountable place to take that objection.

These safeguards do not amount to hostility toward innovation. On the contrary, they identify the conditions under which innovation can be integrated without hollowing out the legal subject. The choice is not between technology and autonomy. It is between technologies embedded in procedures that respect contestable

[16] Regulation (EU) 2016/679 of the European Parliament; Council General Data Protection Regulation of 27 April 2016.

[17] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act).

agency and technologies embedded in procedures that quietly substitute optimisation for judgement.

6. Conclusion

The current regulatory and ethical discussion on AI often asks whether humans remain 'in the loop'. The two case studies analysed in this article show why that formula is insufficient. A human may remain formally in the loop while the practical meaning of choice is nonetheless thinned out by design, workflow and institutional pressure. In the treatment-refusal case, the danger lies in converting statistical recommendation into a new baseline of clinical reasonableness that marginalises competent refusal. In the home-monitoring case, the danger lies in letting safety rationales gradually redefine home as a site of continuous behavioural supervision.

The deeper lesson is that autonomy in technologically mediated care is a procedural achievement. It depends on how information is framed, how reasons are recorded, how dissent is treated, whether consent is revisited when systems expand, and whether professionals are genuinely permitted to depart from the machine's pathway. Legal and administrative analysis should therefore resist the temptation to measure legitimacy by accuracy alone. A highly accurate system can still produce a legally and morally defective practice if it reorganises agency in the wrong way.

For a journal concerned with law and administration, that conclusion has a practical consequence. The task is not merely to classify AI as permissible or impermissible in the abstract. The task is to govern concrete decision environments. When practical case analysis is taken seriously, the relevant question becomes clearer: not whether AI can assist care, but whether institutions are willing to design assistance in a way that leaves the person whose life is at stake recognisable as an author, not merely an input.

References

Appelbaum P.S., *Assessment of Patients' Competence to Consent to Treatment*, "New England Journal of Medicine" 2007, vol. 357(18), pp. 1834-1840.

Foundas M., *Dignity of risk in residential aged care: a call to reframe understandings of risk*, "Medical Journal of Australia" 2025, vol. 223(4), pp. 187-188.

Goddard K., Roudsari A., Wyatt J.C., *Automation bias - a hidden issue for clinical decision support system use*, "Studies in Health Technology and Informatics" 2011, vol. 164, pp. 17-22.

Joseph J., *Designing for dignity: ethics of AI surveillance in older adult care*, "Frontiers in Digital Health" 2025, vol. 7, article 1643238.

O'Neill O., *Autonomy and Trust in Bioethics*, Cambridge 2002.

Parasuraman R., Riley V., *Humans and Automation: Use, Misuse, Disuse, Abuse*, "Human Factors" 1997, vol. 39(2), pp. 230-253.

Perske R., *The Dignity of Risk and the Mentally Retarded*, "Mental Retardation" 1972, vol. 10(1), pp. 24-27.
Raz J., *Legal Rights*, Oxford Journal of Legal Studies 1984, vol. 4(1), pp. 1-21.
Skitka L.J., Mosier K.L., Burdick M., *Does automation bias decision-making?* "International Journal of Human-Computer Studies" 1999, vol. 51(5), pp. 991-1006.
Stauch M., *Comment on Re B (Adult: Refusal of Medical Treatment) [2002] 2 All England Reports 449,* "Journal of Medical Ethics" 2002, vol. 28(4), pp. 232-233.
Thaler R.H., Sunstein C.R., *Libertarian Paternalism*, "American Economic Review" 2003, vol. 93(2), pp. 175-179.
Yeung K., *'Hypernudg": Big Data as a mode of regulation by design,* "Information, Communication & Society" 2017, vol. 20(1), pp. 118-136.