jet INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 3, PP. 1–8
Manuscript received May 17, 2026; revised July 2026. doi: 10.24425/ijet.2026.1727

# Overview of Solutions in Quantum Information Technologies, part 2

Igor Woźniak, Mateusz Jangas, Mateusz Piatek, Karol Franczuk, Piotr Liszewski, Jakub Boruc, Łukasz Siemionek and Ryszard S. Romaniuk

***Abstract*—The hardware available today does not support general-purpose quantum computation in medical settings, but specific subtasks such as sensing, photon-resolved detection, or image reconstruction can already be handled by quantum components embedded in classical systems. This article surveys recent developments across several branches of medical applications, with a complementary look at an analogous trend in machine learning. Nitrogen-vacancy centers in nanodiamonds enable nanoscale thermometry, high-resolution magnetocardiography, and real-time monitoring of free radicals inside living cells. In radiological imaging, photon-counting computed tomography is already entering routine clinical use, while QUBO-based image reconstruction on quantum annealers and gate-model processors points toward dose reduction in low-dose CT. Quantum-enhanced positron emission tomography exploits polarization correlations of entangled annihilation photons to suppress background events and opens a path toward positronium-based biomarkers such as tissue hypoxia. Nuclear magnetic resonance ensemble computing is included to illustrate the scalability limits of fully quantum systems. The same hybrid pattern is then identified outside medicine, in efficient fine-tuning of large language models, where localized quantum modules reduce parameter counts without degrading task accuracy. The reviewed work differs in technical maturity, from systems already deployed clinically to proof-of-principle experiments such as X-ray spontaneous parametric down-conversion. Across this range, the quantum component handles one well-defined stage of an otherwise classical system.**

***Keywords*—quantum sensing, nitrogen-vacancy centers, photon-counting computed tomography, quantum image reconstruction, quantum-enhanced PET, quantum machine learning, hybrid quantum-classical systems**

## I. Introduction

THE promise of quantum computation has long been associated with fault-tolerant universal machines with early popular attention focused on goals such as breaking modern cryptography or simulating molecular systems for drug design. In practice, the hardware available today falls short of this vision: qubit counts are modest, gate fidelities are imperfect, and decoherence times limit the depth of executable circuits. This gap between long-term ambition and present-day capability has shifted attention toward a different mode of integration, in which quantum components do not displace classical infrastructure but are inserted at specific points of an otherwise classical pipeline. The classical system handles the parts where it is mature, scalable, and well understood, while a small quantum module is allocated to a narrowly defined subtask that benefits from access to a Hilbert space of high dimension or from an intrinsically quantum measurement. This pattern has emerged most visibly in medical diagnostics, where the demands on accuracy, dose efficiency, and signal extraction are severe and where even modest hardware can deliver clinically meaningful gains. Nitrogen-vacancy (NV) centers in nanodiamonds turn the spin of a single defect into an optically read probe sensitive to temperature, magnetic fields, and free radicals at the scale of a single cell, supporting applications from intracellular thermometry to magnetocardiography of the beating heart. In radiological imaging, photon-counting detectors exploit the discrete nature of X-ray photons to recover spectral information that energy-integrating systems do not preserve, while reconstruction algorithms cast as Quadratic Unconstrained Binary Optimization (QUBO) problems on quantum annealers and gate-model processors offer a route to lower-dose computed tomography. Positron Emission Tomography (PET) is being extended by polarization-resolved detection of entangled annihilation photon pairs, which suppresses background coincidences and exposes positronium lifetime as a new biomarker of tissue oxygenation. A complementary instance of the same paradigm appears outside medicine, in the efficient fine-tuning of large language models (LLMs), where hybrid quantum-classical adapters replace small but expressive components of the network and reduce parameter counts by significant margins. Against these developments, ensemble quantum computation by nuclear magnetic resonance is reviewed as an early platform whose scaling behavior illustrates why the hybrid approach has prevailed over fully quantum substitutes. Together, the surveyed work spans technologies already in clinical use, systems running on contemporary superconducting hardware, and proof-of-principle experiments that remain confined to the laboratory, providing a current view of where quantum methods plausibly contribute today and where their role is still being established.

The authors are Students at Warsaw University of Technology, Poland (corresponding author e-mail: igor.wozniak.stud@pw.edu.pl). Authors of successive chapters are:
II – M.J., III – K.F., IV – M.P., V – P.L., VI – J.B., Ł.S, VII - I.W.

## II. Quantum Sensors in Medicine

Quantum sensing represents a change in the approach to medical diagnostics, moving away from large-scale, averaged measurements toward the observation of individual quantum states. In classical sensing, the focus lies on macroscopic properties, while in quantum systems the key role is played by features such as electron spin. These systems are highly sensitive to even very small disturbances in their environment, including slight variations in magnetic and electric fields. Such disturbances alter the energy levels of the system, and although these changes are not observed directly, they can be detected through shifts in resonance frequency or optical response. This makes it possible to measure phenomena that remain beyond the reach of classical instruments, while also achieving very high spatial resolution at the nanoscale.

### A. NV Centers as a Sensing Platform

One of the most widely used platforms in biological applications is the nanodiamond with engineered structural defects known as NV centers. This defect forms when a nitrogen atom is located next to a missing carbon atom in the diamond lattice. As a result, a stable site is created that can trap an electron, whose spin behaves like a highly sensitive local probe. An important advantage of this system is that its quantum state can be read optically. When illuminated with a laser, the emitted fluorescence depends on the spin configuration, which allows for straightforward detection. At the same time, diamond is chemically inert and mechanically stable, which means these sensors can operate close to living tissue without causing toxicity or being significantly affected by the surrounding biochemical environment.

### B. Intracellular Thermometry

A particularly important application of this technology is the measurement of temperature inside living cells without damaging them [1]. Traditional fluorescent probes often produce unreliable results because their signals depend on factors such as pH, which interferes with accurate temperature readings. Nanodiamond sensors do not suffer from this limitation, allowing temperature to be measured with a precision better than 0.1°C.

In experimental conditions, gold nanoparticles can be introduced into a cell and used as small heat sources controlled by a laser. Quantum sensors then track how heat spreads within the cell. The results show that temperature differences of several degrees can occur over distances of only a few micrometers while the cell remains stable. By increasing the laser power, it is possible to intentionally reach a point at which the cell dies. This demonstrates the potential for applications in cancer treatment, where it may become possible to selectively destroy cancer cells while monitoring temperature in real time to avoid damaging nearby healthy tissue.

### C. Magnetocardiography

Quantum sensing also offers new possibilities in cardiac diagnostics by enabling more precise mapping of electrical activity. The heart contracts due to electrical currents, and each current generates a magnetic field. Existing technologies such as optically pumped magnetometers provide measurements with limited spatial resolution, which makes it difficult to determine the exact origin of signals. In contrast, sensors based on NV centers allow for much more detailed mapping [2]. By scanning the magnetic field across the heart, it is possible to reconstruct a high-resolution image of electrical activity. Studies on animal models have shown that these sensors can detect signals from specific structures such as Purkinje fibers, which are often not clearly visible using classical methods. In addition, the fast response time of these sensors allows observation of how electrical impulses move through the heart in real time, typically within tens of milliseconds.

### D. Monitoring Cellular Metabolic Activity

In addition to temperature and magnetic field measurements, nanodiamond sensors can be used to study biochemical processes inside cells. One example is the detection of free radicals, which are reactive molecules associated with cellular damage and disease [3]. By placing sensors near mitochondria, researchers can monitor radical production in real time by measuring the T1 relaxation time, which describes how quickly the system returns to equilibrium. When a cell is exposed to harmful substances, this relaxation time decreases, indicating an increase in radical concentration. When antioxidants are introduced, the relaxation time returns toward its original value, which shows that the process can be controlled and monitored. This makes it possible to evaluate the effectiveness of treatments directly at the level of a single cell.

Overall, quantum sensing reflects a broader shift in medicine toward more precise and quantitative analysis at the molecular level. In areas such as rehabilitation, this approach makes it possible to assess tissue recovery based on actual cellular processes rather than only external observation. As a result, it becomes easier to determine whether a patient is ready for further physical effort or requires additional recovery time. This type of measurement can support better evaluation of both therapy and medication, helping to ensure that the healing process is progressing correctly at the most fundamental level.

## III. Ensemble Quantum Computing by Nuclear Magnetic Resonance

The phenomenon of Nuclear Magnetic Resonance (NMR), while predominantly known for its macroscopic applications in medical imaging and chemical analysis, is governed strictly by the quantum mechanical properties of atomic nuclei. The nucleus of an atom consists of protons and neutrons, which are subatomic particles that each possess a quantum spin of $1/2$. Because they naturally pair up with oppositely directed spins, a net spin greater than zero emerges only in nuclei containing an odd number of nucleons. Common examples utilized in quantum computing include the hydrogen isotope ($^{1}H$), carbon-13 ($^{13}C$), and fluorine-19 ($^{19}F$).

When a liquid sample containing these NMR-active nuclei is placed within a strong external magnetic field (denoted as $B_0$), the magnetic dipole moments of the nuclei tend to align with the field lines. However, due to the conservation of angular momentum inherent to the spinning particle, this alignment is not static. Instead, the axis of the spin precesses around the magnetic field lines. This precession occurs at a strictly defined rate known as the Larmor frequency ($f_0$), which is directly proportional to the external magnetic field strength $B_0$ and the gyromagnetic ratio ($\gamma$), a constant unique to each specific element. For instance, in a standard high-field NMR spectrometer operating at 11.7 Tesla, the resonant frequency for $^1$H is approximately 500 MHz, for $^{13}$C it is 126 MHz, and for $^{19}$F it is 470 MHz. This spectral separation is crucial, as it allows researchers to individually address different qubits within the same molecule without causing cross-talk interference.

### A. NMR Qubits and Quantum Gates

In an NMR-based quantum architecture, the discrete spin states of the atomic nuclei function as the qubits. The quantum state $|0\rangle$ corresponds to a nuclear spin oriented parallel to the external magnetic field $B_0$, whereas the state $|1\rangle$ corresponds to a spin oriented anti-parallel to the field. To perform quantum computation, one must be able to manipulate these states with high fidelity. Single-qubit logic gates, such as the Pauli-X (NOT) gate, are executed by applying carefully calibrated radio-frequency (RF) pulses. These RF pulses, which constitute a secondary magnetic field ($B_1$) oscillating at the target nucleus's specific Larmor frequency, rotate the spin state by precise angles, such as a $180^\circ$ flip. Multi-qubit gates, such as the Controlled-NOT (CNOT) gate, are more complex. They are achieved by exploiting the natural spin-electron-spin coupling, widely known in chemistry as J-coupling. Because nuclei within a molecule interact with one another through their shared chemical bonds, the quantum state of one nucleus slightly shifts the resonant frequency of its neighbor. This allows for the execution of conditional logic operations, where a target spin is flipped only if the control spin is in a specific state.

The viability of any quantum computer heavily relies on the coherence times of its qubits. NMR systems are characterized by two primary relaxation times. The longitudinal relaxation time ($T_1$) dictates how long it takes for the magnetization vector to return to thermal equilibrium along the $B_0$ axis. The transverse relaxation time ($T_2$), or spin-spin relaxation, dictates the decay of the magnetization in the perpendicular plane. In liquid samples, $T_2$ times are remarkably long-ranging from hundreds of milliseconds to several seconds. This extended coherence window allows researchers to execute hundreds or even thousands of quantum gate operations before the quantum information is lost to decoherence.

### B. The Pseudo-Pure State Solution

A fundamental limitation of classical NMR technology is its inability to detect the faint magnetic signal of a single isolated atom. To overcome this, NMR quantum computing operates as an Ensemble Quantum Computer (EQC). Instead of measuring a single molecule, the hardware measures the averaged macroscopic magnetic signal emanating from an ensemble of approximately $10^{23}$ identical molecules suspended in a liquid solvent. Each molecule acts as an independent, parallel quantum computer. However, initializing this ensemble poses a severe thermodynamic challenge. To start a quantum computation, the system should ideally be initialized in a pure ground state (e.g., $|00...0\rangle$). Achieving a true pure state in an ensemble of nuclear spins would require cooling the sample to cryogenic temperatures near $10^{-3}$ K. Operating at room temperature means the ensemble exists in a highly mixed state dominated by thermal noise. This barrier was overcome by the landmark development of the pseudo-pure state, thoroughly detailed by Cory, Fahmy, and Havel in 1997 [4]. The researchers devised a method to manipulate mixed thermal populations using precise RF pulses such that the observable output signal mathematically mimics a system initialized in a perfectly pure state. Achieving this requires sophisticated averaging techniques such as spatial averaging, which utilizes magnetic field gradients across the sample tube, causing spins in different physical locations to precess at varying velocities, effectively washing out unwanted signal components. Temporal averaging involves running the exact same quantum algorithm multiple times, but with systematically permuted initial pulse sequences. When the classical computer sums the results of these repeated experiments, the thermal background noise mathematically cancels out, leaving only the amplified signal representing the pure quantum state.

### C. Milestone Implementations

Because of the high degree of control afforded by RF pulses and the exceptionally long $T_2$ coherence times, NMR provided the ideal early testbed for proving quantum algorithms in physical hardware. In 1998, Nielsen, Knill, and Laflamme [5] utilized NMR to achieve complete quantum teleportation. Quantum teleportation involves transferring a quantum state from one location to another without physically moving the particle, relying instead on quantum entanglement. In their experiment, the researchers successfully teleported a spin state from a carbon nucleus to a hydrogen nucleus within a single molecule of trichloroethylene. A few years later, in 2001, Vandersypen *et al.* [6] reported the first experimental realization of Shor's quantum factoring algorithm. Utilizing a custom-synthesized molecule that provided 5 fluorine ($^{19}$F) nuclei and 2 carbon ($^{13}$C) nuclei to act as a 7-qubit quantum register, the ensemble correctly executed the algorithm to determine that the prime factors of 15 are 3 and 5. This remains one of the most famous early demonstrations of applied quantum computing.

### D. The Scalability Bottleneck

Despite its historical importance, NMR quantum computing suffers from a fundamental scalability limit that prevents it from competing with modern architectures like superconducting loops or trapped ions. While engineering sufficiently large, stable molecules with uniquely addressable Larmor frequencies is synthetically difficult, it is not the ultimate bottleneck. The fatal flaw of NMR QC is inextricably linked to the very technique that made it possible: the pseudo-pure state. Because the pseudo-pure state is extracted from a mixed thermal background, the strength of the useful signal diminishes exponentially as the number of qubits increases. Specifically, the signal strength scales proportionally to $n/2^n$, where $n$ is the number of qubits. This limitation was demonstrated in 2006 by Negrevergne *et al.* [7], who successfully benchmarked quantum control methods on a 12-qubit system. At this scale, the usable pseudo-pure signal became so weak that it required exceptionally sensitive measurement coils and over 4,000 temporal repetitions simply to distinguish the algorithm's output from the background thermal noise. Because the signal drops by half with every added qubit, scaling an NMR system to the qubit threshold required for practical quantum supremacy remains beyond current physical limits.

## IV. Quantum Imaging Techniques in X-Ray Imaging

Quantum imaging in X-ray systems combines classical radiographic imaging, quantum optics, and advanced information processing. Its main goal is to improve diagnostic image quality while reducing radiation dose. In practice, this means optimizing the relationship between signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), spatial resolution, and clinical detectability. Two main categories of methods are distinguished. The first one is *quantum-enabled*, where the quantum nature lies in the radiation itself. These methods use entanglement, heralding, coincidence detection, and non-classical photon correlations. The second group is *quantum-inspired*, where the radiation source may remain classical, but the advantage comes from photon-counting detectors, spectral selection, computational reconstruction, and inverse problem solving.

### A. Coincidence Detection, Heralding, and SPDC

Coincidence detection means that two detection events are treated as related because they occur within a very short time window. This is particularly important when photon pairs are generated through spontaneous parametric down-conversion (SPDC), where one high-energy photon is converted into two lower-energy correlated photons. Heralding is a direct consequence of this process. Detection of one photon in the reference arm indicates that its partner is very likely present in the object arm. This allows conditional selection of useful events and rejection of background noise, improving low-dose imaging performance. Entanglement is a stronger quantum resource, where the joint state of two photons cannot be described independently. Although such effects are well known in visible-light quantum optics, X-ray SPDC remains experimentally difficult due to low efficiency, photoionization losses, and strict coincidence requirements. Therefore, these methods are currently mainly proof-of-principle experiments.

### B. Photon-Counting Computed Tomography

Photon-Counting Computed Tomography (PCCT) is the most mature quantum-inspired X-ray imaging technique.

Conventional Computed Tomography (CT) uses Energy-Integrating Detectors (EID), which measure the total deposited energy over time. In contrast, photon-counting detectors register each photon separately and estimate its energy. This enables spectral imaging, better material decomposition, improved soft-tissue contrast, higher spatial resolution, and reduction of beam-hardening artifacts. The quantum aspect of PCCT does not come from entanglement, but from direct photon-resolved detection. Each event corresponds to one absorbed X-ray photon, making PCCT the most clinically relevant example of quantum-inspired imaging already entering routine medical practice [8].

### C. Ghost Imaging and Computational Ghost Imaging

Ghost imaging (GI) reconstructs an image using correlations rather than direct pixel-by-pixel detection. Instead of a spatial detector, the object is measured with a bucket detector, which records only the total transmitted intensity. Spatial information is obtained from a second reference arm, where the illumination pattern is known. The image is reconstructed from correlations between the bucket signal and the reference signal. Computational ghost imaging (CGI) removes the physical reference detector and uses predefined illumination patterns. Only the bucket signal is measured, while the image is reconstructed numerically. This shifts system complexity from hardware to computation and allows single-pixel X-ray imaging. Although GI and CGI are often associated with low-dose imaging, reduced instantaneous signal does not automatically mean lower total dose. The final performance depends on the number of measurements, noise level, reconstruction model, and preservation of diagnostic quality [9], [10].

### D. SNR, CNR, and Diagnostic Utility

Image quality is commonly evaluated using SNR and CNR. In classical imaging, SNR is fundamentally limited by shot noise, which follows Poisson statistics. Improving image quality therefore requires increasing the number of detected photons and, consequently, radiation dose. CNR is often even more important clinically, because it determines whether healthy tissue and pathological structures can be reliably distinguished. A high-resolution image may still be diagnostically useless if contrast is too low compared to noise. Quantum and quantum-inspired methods attempt to improve this trade-off in two ways: by reducing noise through photon correlations as in sub-shot-noise imaging, or by changing the measurement architecture, as in GI and CGI. However, no method guarantees automatic dose reduction. True advantage exists only when equivalent diagnostic quality is preserved.

The near future of X-ray imaging belongs mainly to quantum-inspired methods, especially PCCT, which already offers real clinical benefits. Quantum-enabled methods such as X-ray SPDC, coincidence imaging, and heralded detection remain important research directions but are still far from routine applications. Quantum imaging should therefore be understood not only as the use of entangled photons, but as a broader integration of radiation physics, photon-counting detection, and computational reconstruction. The future of X-ray imaging depends increasingly on how efficiently information can be extracted from every detected photon.

## V. Quantum Optimization in CT Image Reconstruction

CT has undergone a fundamental transition from EID to PCCT. The central focus is the application of the QUBO model in the image reconstruction process. Traditional analytical algorithms, such as Filtered Back Projection (FBP), exhibit limitations in handling noise and incomplete sinogram data. The QUBO model allows reconstruction to be formulated as an energy minimization problem, where each pixel is represented by a system of qubits. This approach enables higher fidelity images, artifact resistance [11], and simultaneous tissue segmentation directly from raw data [12]. The use of quantum annealers and hybrid algorithms may significantly reduce radiation dose required in Low-Dose CT (LDCT), while maintaining diagnostic value.

### A. From Traditional to Quantum Detection

CT is the foundation of modern diagnostic imaging, utilizing X-rays to generate cross-sectional maps of radiological density. This phenomenon is based on attenuation of the photon beam:

$$\mu_E t = -\frac{1}{d}\ln\left(\frac{I}{I_0}\right)\ [cm^{-1}]$$

where $I$ and $I_0$ are incident and transmitted beam intensities, and $d$ is material thickness.

In traditional EID architecture, detection is a two-stage process: a scintillator converts X-ray photons into visible light, then a photodiode generates an electrical signal. This causes information loss regarding photon energy and introduces electronic noise. The breakthrough is photon-counting technology using semiconductors such as cadmium telluride (CdTe). These detectors operate under high voltage around 1000 V and with pixelated anodes directly convert X-ray photons into electron-hole pairs, eliminating background noise and enabling spectral analysis of every event.

This direct conversion is not without practical limitations. At high X-ray flux, two photons may arrive within the detector dead time, so that their pulses overlap and partially merge, an effect known as pile-up, which causes count loss and biased energy estimation. A second issue, charge sharing, arises when a photon interacts close to a pixel boundary and the resulting charge cloud is split between adjacent pixels, degrading both spatial resolution and the accuracy of energy binning. Faster ASICs, anti-coincidence logic between neighboring pixels, and model-based correction algorithms are used to mitigate both effects, and remain an active area of detector engineering [13].

### B. Mathematical Foundations of Quantum Image Reconstruction

In the quantum model, a CT image of size $n \times n$ is no longer a matrix of scalar Hounsfield units but a quantum state mapped onto processor architecture. Each pixel $I_{ij}$ is represented as a linear combination of qubits and binary weights:

$$I_{ij} = \sum_{k=0}^{m} 2^k q_{ij}^k$$

where $q_{ij}^k \in \{0,1\}$ denotes the state of the $k$-th qubit assigned to a pixel.

The relationship between the qubit-domain image and the sinogram – the discrete measurement data – is defined by the discrete Radon transform:

$$IP(\theta, s) = \sum_{i,j} c_{ij} I'_{ij}$$

where $c_{ij}$ represents the geometric contribution of pixel $I'_{ij}$ to a projection line.

### C. QUBO Model and System Energy Minimization

Reconstruction as a QUBO problem seeks a qubit configuration minimizing the difference between model and measured reality:

$$F = \sum_{\theta}\sum_{s}\left(IP(\theta, s) - P(\theta, s)\right)^2$$

Using binary identity $q^2 = q$, the error function becomes a QUBO matrix containing linear diagonal terms and quadratic off-diagonal qubit interactions. Constant terms are omitted because they do not affect the global minimum. Calculations are performed on gate-model processors as those offered by IBM using QAOA, or on annealers such as D-Wave Advantage. Hybrid solvers decompose large QM matrices into smaller quantum-solvable subproblems.

### D. Advantages of the QUBO Model over Classical Methods

- **Global vs. Local Optimization:** Classical iterative methods (SART, MLEM) may become trapped in local minima. Quantum tunneling improves search for the global minimum.
- **Artifact Resistance:** Simultaneous optimization of the entire sinogram mitigates systematic errors such as ring or motion artifacts.
- **Exceeding the Projection Barrier:** Accurate reconstruction may be possible with fewer projections than $n_x$ for a sample size $n_x \times n_x$, which is essential for Low-Dose CT.

### E. Simultaneous Segmentation and Post-processing

This innovation blurs the line between reconstruction and segmentation. In classical approaches, segmentation is secondary and inherits image-creation errors. Using attenuation coefficients $\alpha = \mu/\rho$, the search space can be restricted to discrete tissue classes such as bone, muscle, and blood, which are far fewer than grayscale levels (1024–4096), thus significantly decreasing qubit-per-pixel requirements. This enables direct extraction of anatomical structures from the sinogram.

Quantum CT optimization using the QUBO model is a foundation for future precision 3D diagnostics. Although hardware limits remain significant, with D-Wave Advantage offering about 180 fully connected logical variables, limiting direct solvers to $13 \times 13$ pixel images, hybrid solvers already allow larger matrices. The QUBO model increases artifact resistance and enables major radiation dose reduction while maintaining high fidelity.

## VI. Integrating Quantum Entanglement into Positron Emission Tomography

While CT focuses on anatomical structures using external X-ray sources, PET visualizes biological functions by detecting radiation from internal radioactive isotopes. Current PET technology relies on the detection of two 511 keV photons produced during the annihilation of a positron and an electron [14]. However, conventional PET systems discard significant diagnostic information by ignoring the quantum nature of these photons: classical PET is quantum-blind, measuring only the time and direction of photon arrival while completely ignoring the polarization states of the emitted photons [14]. This omission leads to the registration of false Lines of Response (LOR) caused by random coincidences and Compton scattering, degrading image resolution and contrast. Quantum-enhanced PET (QE-PET) addresses this limitation by leveraging the entanglement of annihilation photon pairs and utilizing polarization correlations as a quantum filter, with recent research demonstrating the potential for higher signal-to-noise ratios (SNR), radiation dose reduction, and new biomarkers such as tissue hypoxia [15].

### A. The Physics of Quantum Entanglement in PET

Quantum entanglement is a phenomenon where two particles form a non-local quantum system, such that the state of one particle instantaneously determines the state of the other. In the context of PET, photons originating from the annihilation of para-positronium (p-Ps) emerge in an entangled state with strictly orthogonal linear polarization correlation. This state can be formally described using Bell states, where the polarization planes of the two photons are intrinsically linked. However, the biological environment introduces complexity. Phenomena such as pick-off annihilation, where a positron annihilates with an external electron from the surrounding tissue, can disrupt this entanglement. Experimental evidence shows that photon pairs in matter exhibit non-maximal entanglement, depending on the annihilation mechanism [14]. While theoretical vacuum annihilation yields a maximal entanglement coefficient of $R \approx 2.85$, experimental measurements in tissue-like conditions show a reduction due to environmental interactions.

### B. Experimental Breakthroughs: J-PET and Polarimetric Detection

The shift toward QE-PET has been accelerated by the development of innovative hardware. The J-PET (Jagiellonian PET) scanner represents a plastic scintillator-based system capable of measuring photon polarization through the analysis of Compton scattering angles [14]. Using such systems, it has been demonstrated that the relative angle between polarization planes carries information about the degree of entanglement, enabling discrimination between true and background events. Simultaneously, positronium-based imaging has emerged as a complementary paradigm. Recent studies demonstrate that positronium lifetime imaging is feasible in vivo and provides additional physiological information beyond standard PET, including sensitivity to oxygen concentration and tissue microstructure [15], [16].

### C. QE-PET and Clinical Benefits

The application of Quantum Entanglement Filtering (QEF) offers several transformative advantages for clinical practice:

- **Superior SNR:** Entanglement-based selection allows rejection of scattered and random events, since only true annihilation pairs preserve polarization correlations [14].
- **Dose Reduction:** Improved event selection efficiency may reduce the required radiotracer activity, analogous to improvements seen in advanced PET methodologies.
- **Quantum Biopsy:** Positronium lifetime is sensitive to tissue oxygenation and molecular structure, enabling non-invasive assessment of hypoxia and early pathological changes [15], [16].

### D. Technological Challenges and the Roadmap

Despite the potential, several barriers remain. Extracting quantum information requires high-precision polarimetric detectors capable of resolving Compton scattering kinematics. Furthermore, reconstructing images that incorporate polarization or positronium observables requires advanced reconstruction frameworks. From a clinical perspective, positronium imaging is still in early translation stages, with initial in vivo human studies demonstrating feasibility but requiring further validation and standardization before routine use [15].

The future of medical diagnostics is increasingly quantum. By transitioning from classical coincidence detection to the exploitation of photon entanglement, QE-PET promises to redefine the limits of sensitivity and precision in oncology. As demonstrated by recent experimental work, both entanglement-sensitive PET and positronium imaging are progressing toward clinically relevant tools capable of delivering fundamentally new diagnostic information [14], [15].

## VII. Quantum Extension of LLMs for Efficiency

The computational cost of training and adapting LLMs has grown into one of the biggest bottlenecks of modern natural language processing. Full fine-tuning of Transformers with billions of parameters is excessively expensive for most research groups, which has motivated an ongoing search for alternative methods. Although full-scale quantum computing for such workloads remains beyond the reach of current technology, this has not discouraged attempts to replace individual components of classical models with their quantum counterparts. Superposition and entanglement can, in principle, grant access to an exponentially large Hilbert space with fewer trainable parameters and, in some cases, even better task performance than a purely classical infrastructure.

### A. Quantum Tensor Hybrid Adaptation

One of the most prominent classical strategies for reducing fine-tuning cost is Low-Rank Adaptation (LoRA) [17], which constrains weight updates to a low-rank subspace of the original parameter matrix. While effective, the expressive capacity of LoRA is bounded by the chosen rank, and in complex, high-rank-dependent tasks it may underfit, converge slowly or exhibit sensitivity to the rank hyper-parameter [17]. The Quantum Tensor Hybrid Adaptation (QTHA) framework proposed by Kong *et al.* [18] directly addresses this limitation by replacing the classical low-rank adapter with a hybrid module that combines a Matrix Product Operator (MPO) tensor network and a Quantum Neural Network (QNN). The MPO decomposes the original weight matrix $W \in \mathbb{R}^{N_y \times N_x}$ into a product of local tensors with bond dimension $D$, reducing the parameter count from $N_x N_y$ to a sum governed by $D$ and the chosen factorization. The QNN, implemented as a parameterized quantum circuit with $R_Y$ angle embeddings and controlled-$R_Z$ entangling gates, processes the intermediate representation and produces a classical vector through Pauli-$Z$ expectation values. The two branches are then combined linearly:

$$\tilde{O} = W_q O_q + W_c O_c,$$

where $O_q$ and $O_c$ are the quantum and classical branch outputs. These outputs encode complementary information: the MPO captures abrupt, non-harmonic components, whereas the QNN efficiently fits truncated Fourier series [18], [19], so their combination enlarges the class of learnable functions, while keeping the parameter count low.

Empirically, QTHA reduces the number of trainable parameters by roughly 76% compared to a rank-matched LoRA baseline on DeepSeek-R1-Distill-Qwen-7B and Qwen2-7B-Instruct, while lowering the training loss by up to 17% and improving standard text-generation metrics by a similar margin across three benchmarks. Convergence is also faster, with the loss reaching its plateau about 20% earlier than with LoRA. The parameter reduction is the headline number, but in practice the accelerated convergence may be more important, since training cost scales with steps as well as with parameter count. The inference itself is performed on an actual superconducting quantum computer, *Origin Wukong*, and the authors report that hardware shot noise does not destroy the attained accuracy. In several cases the real-device results slightly outperform the noiseless simulator, an effect attributed to implicit regularisation by quantum noise.

### B. Complementary Hybrid Architectures

The findings of QTHA gain additional support from two complementary strategies that target different points of the LLM pipeline. Kim *et al.* [20] adopt a conservative approach: the body of a Sentence Transformer (SetFit) is kept frozen and a parameterized quantum circuit is attached as the classification head for sentiment analysis on SST-2. Despite a very small parameter budget (353–914 trainable parameters across 10–18 qubits), their best hybrid model attains 92.70% accuracy, a 3.14% gain over a size-matched Support Vector Classifier, falling slightly short of the fully fine-tuned classical upper bound of 93.62%. Their energy analysis further estimates a cross-over at 46 qubits beyond which QPU inference consumes less energy than the equivalent simulation on an H100 GPU, providing empirical evidence for the efficiency.

A more aggressive integration is explored in Hybrid Quantum Transformer (HyQuT) [21], where variational quantum circuits replace entire linear projections inside the Transformer itself, either the gate projection of the feed-forward block at 8M parameters or the query projection $W_Q$ of the self-attention block at 150M. With only 10 qubits and about 80 gates, HyQuT removes roughly 10% of the classical parameters while retaining stable convergence and coherent text generation after pre-training on a 1.6 GB Chinese corpus. An ablation study shows, however, that replacing all four attention matrices or the full Feed-Forward Network prevents convergence, indicating that current quantum modules are best inserted surgically at a single, carefully chosen architectural point.

Comparing the three approaches reveals a consistent pattern: hybrid quantum-classical components can reduce parameter counts and, in some regimes, energy usage without degrading task accuracy, provided they occupy a well-defined, limited role within the network. QTHA is the most mature of the three in the fine-tuning context, as it directly generalises the widely adopted LoRA adapter and validates its results on physical quantum hardware. The lightweight quantum head of Kim *et al.* and the deeper but narrowly scoped replacement in HyQuT confirm the same underlying principle from opposite ends of the intervention spectrum. The remaining obstacles such as limited qubit counts, the overhead of gradient estimation for quantum parameters, and the absence of fault-tolerant hardware suggest that, in the near term, quantum extensions will function as compact, strategically inserted modules, rather than as wholesale substitutes for the classical backbone.

## VIII. Conclusion

The technologies reviewed in this article share a common architectural choice: quantum components are deployed as narrowly scoped modules within larger classical systems, rather than as full replacements for them. In medical sensing, NV centers contribute as compact probes embedded in otherwise conventional optical readout chains. In radiological imaging, photon-counting detectors and QUBO-based reconstruction operate inside CT pipelines whose acquisition geometry, calibration, and clinical workflow remain classical. QE-PET adds polarization-resolved detection on top of established coincidence hardware. In machine learning, hybrid adapters such as QTHA and HyQuT modify only specific layers of transformers that are otherwise trained and deployed by standard means. The case of nuclear magnetic resonance ensemble computing reinforces the same conclusion from the opposite direction. When an entire computational system is made quantum, the signal-to-noise penalty scales unfavorably with problem size, and the platform settles into a role of historical importance rather than practical scaling.

The reviewed methods occupy very different points on the maturity spectrum. PCCT is already part of clinical practice, NV-based sensing has reached reproducible in vivo demonstrations, and quantum-enhanced fine-tuning of LLMs has been validated on superconducting hardware. At the other end, X-ray spontaneous parametric down-conversion, large-scale QUBO reconstruction, and clinical positronium imaging remain at proof-of-principle or early-translation stages. The term quantum therefore covers a wide range of technological maturity, and the relevant question for each technology is which specific subtask it improves and at what cost in hardware complexity, calibration, and signal extraction. The hybrid pattern suggests that the value of near-term quantum technology lies less in algorithmic supremacy than in providing access to physical resources unavailable to classical systems: single-spin sensitivity, photon-resolved energy spectra, polarization correlations of entangled pairs, and exponentially large parameter spaces accessed through a small number of qubits. Progress in each of the surveyed areas will depend on co-design between the quantum module and the classical infrastructure it augments. It will also require standardization and validation in clinical contexts where applicable, and honest accounting of where a quantum component genuinely outperforms a well-tuned classical alternative.

## References


[1] G. Kucsko, P. C. Maurer, N. Y. Yao, M. Kubo, H. J. Noh, P. K. Lo, H. Park, and M. D. Lukin, "Nanometre-scale thermometry in a living cell," *Nature*, vol. 500, no. 7460, pp. 54–58, 2013.

[2] K. Arai, A. Kuwahata, D. Nishitani, I. Fujisaki, R. Matsuki, Y. Nishio, Z. Xin, X. Cao, Y. Hatano, S. Onoda *et al.*, "Millimetre-scale magnetocardiography of living rats with thoracotomy," *Communications Physics*, vol. 5, no. 1, p. 200, 2022.

[3] L. Nie, A. Nusantara, V. Damle, R. Sharmin, E. Evans, S. Hemelaar, K. Van der Laan, R. Li, F. Perona Martinez, T. Vedelaar *et al.*, "Quantum monitoring of cellular metabolic activities in single mitochondria," *Science advances*, vol. 7, no. 21, p. eabf0573, 2021.

[4] D. G. Cory, A. F. Fahmy, and T. F. Havel, "Ensemble quantum computing by NMR spectroscopy," *Proceedings of the National Academy of Sciences*, vol. 94, no. 5, pp. 1634–1639, 1997.

[5] M. A. Nielsen, E. Knill, and R. Laflamme, "Complete quantum teleportation using nuclear magnetic resonance," *Nature*, vol. 396, no. 6706, pp. 52–55, 1998.

[6] L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang, "Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance," *Nature*, vol. 414, no. 6866, pp. 883–887, 2001.

[7] C. Negrevergne, T. S. Mahesh, C. A. Ryan, M. Ditty, F. Cyr-Racine, W. Power, N. Boulant, T. Havel, D. G. Cory, and R. Laflamme, "Benchmarking quantum control methods on a 12-qubit system," *Physical Review Letters*, vol. 96, no. 17, p. 170501, 2006.

[8] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, 2018.

[9] D. Ceddia and D. M. Paganin, "Random-matrix bases, ghost imaging, and x-ray phase contrast computational ghost imaging," *Physical Review A*, vol. 97, no. 6, p. 062119, 2018.

[10] D. Pelliccia, A. Rack, M. Scheel, V. Cantelli, and D. M. Paganin, "Experimental X-ray ghost imaging," *Physical Review Letters*, vol. 117, no. 11, p. 113902, 2016.

[11] K. Jun, "A highly accurate quantum optimization algorithm for CT image reconstruction based on sinogram patterns," *Scientific Reports*, vol. 13, p. 14407, 2023.

[12] K. Jun and H. Lee, "Quantum optimization algorithms for CT image segmentation from X-ray data," *Scientific Reports*, vol. 15, p. 20649, 2025.

[13] S. S. Hsieh, S. Leng, K. Rajendran, S. Tao, and C. H. McCollough, "Photon counting CT: Clinical applications and future developments," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 4, pp. 441–452, 2021.

[14] P. Moskal, D. Kumar, S. Sharma *et al.*, "Nonmaximal entanglement of photons from positron-electron annihilation demonstrated using a plastic PET scanner," *Science Advances*, vol. 11, no. 18, p. eads3046, 2025.

[15] P. Moskal, A. Bilewicz, M. Das, B. Huang, A. Khreptak, S. Parzych, J. Qi, A. Rominger, R. Seifert, S. Sharma, K. Shi, W. Steinberger, R. Walczak, and E. Stepień, "Positronium imaging: History, current status, and future perspectives," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 9, no. 8, pp. 981–1001, 2025.

[16] D. P. Watts, J. Bordes, J. R. Brown, A. Cherlin, R. Newton, J. Allison, M. Bashkanov, N. Efthimiou, and N. A. Zachariou, "Photon quantum entanglement in the MeV regime and its application in PET imaging," *Nature Communications*, vol. 12, p. 2646, 2021.

[17] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," in *International Conference on Learning Representations*, 2022.

[18] X. Kong, L. Li, Z. Chen, C. Xue, X. Xu, H. Liu, Y. Wu, Y. Fang, H. Fang, K. Chen, Y. Yang, M. Dou, and G. Guo, "Quantum-enhanced llm efficient fine tuning," 2025. [Online]. Available: https://arxiv.org/abs/2503.12790

[19] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Physical Review A*, vol. 103, p. 032430, 2021.

[20] S. H. Kim, J. Mei, C. Girotto, M. Yamada, and M. Roetteler, "Quantum large language model fine-tuning," 2025. [Online]. Available: https://arxiv.org/abs/2504.08732

[21] D. Kong, X. Cui, J. Jin, J. Xu, and D. Wang, "Hybrid quantum transformer for language generation," 2025. [Online]. Available: https://arxiv.org/abs/2511.10653